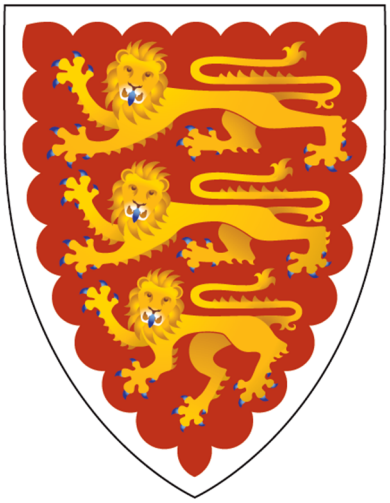
B Cells,

T Cells and Sepsis:

Understanding the Adaptive Immune Response



Lauren Elizabeth Overend

Oriel College

The University of Oxford

Thesis Submitted for Degree of

Doctor of Philosophy

August 2023

‘All you really need to know for the moment is that the universe is a lot more complicated than you might think, even if you start from a position of thinking it's pretty damn complicated in the first place.’

— Douglas Adams

Declaration

This thesis describes work carried out between October 2019 and August 2023, under the supervision of Associate Professor Rachael Bashford-Rogers and Professor Julian Knight at the Wellcome Trust Centre for Human Genetics, The University of Oxford. The work presented in this thesis is my own, unless specifically indicated in the text as a collaboration.

This thesis is **465** pages in length and text has been typeset in 11pt Computer Modern Serif font, with 2.0 line spacing. Figures were generated using BioRender.

Abstract

Sepsis is defined as life-threatening organ dysfunction resulting from a dysregulated host response to infection. Globally, sepsis is responsible for 11 million deaths each year and is the most common reason for admission to adult intensive care units (ICUs) in the UK. Despite its significant impact, there are currently no sepsis specific treatments targeting the host response. Sepsis is associated with high mortality, estimated at 30% for the initial episode, and around 15% of survivors die within the following year. Healthcare-associated or secondary infections and recurrent sepsis are common causes of long-term mortality, suggesting that persistent alterations in adaptive immunity contribute to the detrimental effects of sepsis. A greater understanding of underlying disease mechanisms is paramount for reducing the associated socio-economic burden.

To elucidate the adaptive immune response in sepsis, I have conducted the first comprehensive longitudinal study of the B and T cell receptor (B/TCR) repertoire using high-throughput multiplex PCR and next generation sequencing (NGS). Seventy-two patients with sepsis and twelve matched healthy controls were sampled at three time points, days one, three and five, post-ICU admission to understand the dynamic immune response. Patients were recruited as part of the UK Genomic Advances in Sepsis (GAinS) study.

As part of this thesis, I developed an optimised data dimensional reduction pipeline called repertoire (r) WGCNA which utilises weighted gene correlation network analysis. I demonstrate how rWGCNA can be used to identify a small number of modules which capture the combined effect of several correlated immune repertoire features (*e.g.* diversity and V gene usage). By applying conventional statistical analysis, I used these modules to characterise the alterations to the adaptive immune repertoire during sepsis, effectively distinguishing between health and disease. Additionally, I identified repertoire features with significant associations to clinical variables, Sepsis Response Signatures (SRS), and mortality, focusing on novel biology and generating hypotheses for future research.

Finally, by integrating matched bulk RNA sequencing data from whole blood, I identified genes and pathways associated with various aspects of the adaptive immune response, including clonal expansion and diversification. These findings provide a foundation for future research aimed at understanding the pathophysiology of sepsis.

Acknowledgments

The work presented in this thesis would not have been possible without the guidance and support provided by many exceptional scientists throughout my PhD and career. First and foremost, I would like to thank my supervisors, Associate Professor Rachael Bashford-Rogers and Professor Julian Knight, for their unwavering encouragement and expertise which was particularly helpful during the challenging months of remote work amidst the COVID-19 pandemic. A special mention must go to my post-doctoral colleagues, Dr Felicia Tucci, who helped me find my feet in the lab and patiently listened to the endless trials and tribulations of multiplex PCR optimisation and also Dr Bo Sun whose intellectual discussions pushed me to be a better scientist. On the computational side, I would like to acknowledge Santiago Revale, whose knowledge of bioinformatics pipelines was instrumental in developing my code, else I would still be analysing samples in series with an estimated completion year of 2099. Thanks also to Assistant Professor Justin Whalley, Dr Katie Burnham and Dr Emma Davenport, for their statistical and bioinformatics guidance on the analysis presented in this thesis.

I extend my appreciation to the members of the Bashford-Rogers and Knight research groups for their friendship, feedback and advice. Additionally, I would like to acknowledge the patients and participants of the Genomic Advances in Sepsis study, who made this research possible. I am grateful for the generous funding provided by the DPhil in Genomic Medicine and Statistics program, supported by the Wellcome Trust, which enabled me to fulfil both personal and scientific goals throughout my PhD.

On a personal level I would like to express my biggest thanks to my parents (Helen and Tim Overend) who have always encouraged me to pursue my goals and proof-read every essay, CV and personal statement I sent their way. I would not be writing this today without the support they have shown me every step of the journey and I am very proud to be Dr Overend 2.0. I am also grateful to my friends who are always encouraging and supportive of my dreams – even when I want to start a bakery in the jungle. I would like to finish by dedicating this thesis to my grandparents, Geoff and Jean Overend, who always believed in me.

Lauren E. Overend, August 2023

Table of Contents

Declaration **2**
Abstract **3**
Acknowledgments **4**
Table of Contents **5**
List of Figures **9**
List of Tables **13**
List of Equations **14**
Nomenclature **15**
1 Introduction **17**
1.1 Sepsis 18
1.1.1 Introduction and Disease Burden 18
1.1.2 Clinical Definition 20
1.1.3 Dysregulation of the Inflammatory Response in Sepsis 24
1.1.4 Sepsis Immunosuppression 25
1.1.5 Sepsis Aetiology and Epidemiology 26
1.1.6 Clinical Microbiology 29
1.1.7 Genomic Advances in Sepsis Study 31
1.2 The Adaptive Immune System 32
1.2.1 Pathology in Sepsis 32
1.2.2 Biomarkers in Sepsis 34
1.2.3 Targeted Therapeutics in Sepsis 36
1.2.4 B Cell Biology 37
1.2.5 T Cell Biology 56
1.3 The Adaptive Immune Repertoire 66
1.3.1 Generating Diversity 66
1.3.2 Repertoire Analysis Methods 71
1.4 Chapter 1 Aims 77
2 Materials and Methods **80**
2.1 Sample Selection 80
2.1.1 UK Genomic Advances in Sepsis Study 80
2.1.2 Sepsis Samples 81
2.1.3 Healthy Volunteer Recruitment 83
2.1.4 Sample Collection 84
2.2 BCR/TCR Multiplex PCR Repertoire Amplification 84
2.2.1 Preventing Contamination 84
2.2.2 RNA Extraction 85

2.2.3 Reverse Transcription 85
2.2.4 Multiplex PCR 87
2.2.5 Gel Electrophoresis 93
2.2.6 Library Preparation 93
2.2.7 Sequencing 95
2.3 Bulk RNA Sequencing 95
2.4 Computational Analysis 96
2.4.1 Bulk RNA Sequencing 96
2.4.2 Pre-processing: Bulk BCR/TCR Repertoire Analysis 100
2.4.3 Comparing Isotype usage between Bulk RNA Sequencing and Targeted BCR Repertoire Amplification 111
2.5 Statistical Analysis 115
2.5.1 Repeated Measures Correlation 115
2.5.2 Linear Mixed Effects Models 115
2.5.3 Multiple Testing Correction 118
2.6 B Cell Clonal Dynamics with MRDARCY 119
2.6.1 Method Summary 119
2.6.2 Clonal Dynamics 119
2.6.3 Clonal Overlap 120
2.6.4 BCR Longitudinal Clone Network Generation 121
3 The Development of Computational Methods to Analyse the BCR and TCR Repertoire 123
3.1 Introduction 123
3.1.1 Immune Repertoire Sequencing Analysis 124
3.2 Chapter 3 Aims 128
3.3 Cohort Summary 128
3.4 Data Quality Control Analysis 130
3.4.1 Batch Effect Assessment 130
3.4.2 Primer Validation 135
3.4.3 Index Hopping and Sample Contamination 136
3.4.4 Rarefaction Analysis 140
3.4.5 Quality Control Summary 141
3.5 Method Development: Dimensional Reduction of Repertoire Feature Matrices 142
3.5.1 Introduction and Aims 142
3.5.2 Repertoire Weighted Gene Correlation Network Analysis (rWGCNA) 145
3.5.3 Tensor Decomposition: Sparse Decomposition of Arrays 169
3.6 Discussion 177
3.6.1 Comparison of Dimensional Reduction Methods 177
3.6.2 Unbalanced Design 178

3.6.3 Consensus Module Generation 179
3.6.4 Imputation and Missing Data 180
3.6.5 AIRR Minimum Standards and Package Development 183
3.6.6 VDJ Sequence Representation 183
3.7 Conclusion 185
4 The BCR Repertoire in Sepsis 187
4.1 Introduction 187
4.1.1 The BCR Repertoire in Sepsis 188
4.1.2 Age Related Changes to the BCR Repertoire 191
4.1.3 Sex Related Changes to the BCR Repertoire 192
4.2 Chapter 4 Aims 193
4.3 Results 195
4.3.1 Statistical Analysis of BCR Repertoire Modules 195
4.3.2 Integrative Analysis 212
4.3.3 Interpreting BCR Repertoire Module Biology in the Context of Sepsis 222
4.3.4 Longitudinal B cell Clonal Dynamics 256
4.4 Discussion 270
4.4.1 Unpaired IgH Chain Sequencing 270
4.4.2 Sepsis and Accelerated Immunosenescence 271
4.5 Conclusion 272
5 The TCR Repertoire in Sepsis 274
5.1 Introduction 274
5.1.1 The TCR Repertoire in Sepsis 275
5.1.2 The Effect of Age and Sex on the TCR Repertoire 279
5.2 Chapter 5 Aims 280
5.3 Results 281
5.3.1 Summary of Study Cohort 281
5.3.2 Quality Control Analysis 282
5.3.3 Evaluating the Performance of rWGCNA Analysis 300
5.3.4 Statistical Analysis of TCR Repertoire Modules in Sepsis 303
5.3.5 Characterising TCR Repertoire Modules in Sepsis 333
5.3.6 Integrating TCR and BCR Repertoires 353
5.4 Discussion 356
5.4.1 TCR Repertoire Sequencing: Technical Challenges 356
5.4.2 Sepsis is Characterised by Subtle Alterations to the TCR repertoire 358
5.5 Conclusion 360
6 Discussion and Future Directions 363
6.1 Overall Summary: The Adaptive Immune Repertoire in Sepsis 363

6.2 Discussion and Limitations 365
6.2.1 Adaptive Immune Repertoire Stability in Health 366
6.2.2 Time of Sepsis Onset 367
6.2.3 Infectious Aetiology: Community-Acquired Pneumonia versus Faecal Peritonitis 369
6.2.4 The Dysregulated Immune Response and Choice of Controls 371
6.2.5 Racial Disparity in Sepsis 374
6.3 Future Directions 375
6.3.1 Repertoire Reconstruction from Bulk RNA Sequencing 375
6.3.2 Receptor Specificity 377
6.3.3 Integration of Multi-Omic Data 379
6.4 Conclusion 381
***A* Appendices: Materials and Methods 382**
***B* Appendices: The Development of Computational Methods to Analyse the BCR and TCR Repertoire 390**
***C* Appendices: The BCR Repertoire in Sepsis 398**
***D* Appendices: The TCR Repertoire in Sepsis 421**
References 445

List of Figures

Figure 1. Heterogeneity in sepsis. 20
Figure 2. Sepsis disease trajectory. 23
Figure 3. The dysregulated adaptive immune response in sepsis. 34
Figure 4. Representative structure of the human B cell receptor (BCR) and antibody. 40
Figure 5. B cell maturation process. 48
Figure 6. Class switch recombination network. 56
Figure 7. Representative structure of the human T cell receptor (TCR) complex. 59
Figure 8. T cell development. 62
Figure 9. IgH chain VDJ recombination. 68
Figure 10. High-throughput BCR and TCR repertoire amplification methods. 77
Figure 11. Sepsis patient demographics for adaptive immune repertoire cohorts 1 and 2. 83
Figure 12. In-house BCR repertoire multiplex PCR amplification protocol schematic. 88
Figure 13. In-house TCR repertoire multiplex PCR amplification protocol schematic. 89
Figure 14. Bulk RNA sequencing batch effect assessment. 98
Figure 15. SRS assignment from bulk RNA sequencing gene expression data, using the 'SepstratifieR' R package. 99
Figure 16. In-house bulk BCR/TCR computational analysis pipeline 101
Figure 17. Good correlation between the percentages of the repertoire comprised of each isotype, as output from the targeted BCR repertoire analysis pipeline (y axis) versus the normalised bulk RNA sequencing (x axis) isotype measure. 113
Figure 18. Poor correlation between the percentages of the repertoire comprised of each isotype, as output from the targeted BCR repertoire analysis pipeline (y axis) versus the corresponding immunoglobulin isotype gene expression from bulk RNA sequencing data (x axis). 114
Figure 19. The BCR repertoire in sepsis study schematic. 129
Figure 20. QC summary: BCR repertoire sequencing depth post-filtering, for all samples. 131
Figure 21. QC summary: immunoglobulin isotype usage for the 10 technical replicates (healthy donor JR-1975-1003) included in the BCR repertoire cohort. 134
Figure 22. QC Summary: IGHV/J gene usage for the 10 technical replicates (healthy donor JR-1975-1003) included in the BCR repertoire cohort. 135
Figure 23. QC summary: total IGHV counts in the BCR repertoire data. 136
Figure 24. Schematic depicting computational detection of index hopping in sequencing data, due to the use of combinatorial library adaptors in conjunction with internal PCR barcodes. 137
Figure 25. QC summary: unique BCR VDJ sequence overlap between pairs of samples. 139
Figure 26. QC summary: BCR repertoire rarefaction analysis. 141
Figure 27. Repertoire summary metrics output by the in-house pre-processing pipeline. 142
Figure 28. Custom rWGCNA dimensional reduction pipeline utilising the principles of weighted gene correlation network analysis (WGCNA). 147
Figure 29. BCR repertoire missing data imputation optimisation. 151
Figure 30. BCR repertoire imputation performance summary. 153
Figure 31. Comparing correlation matrices derived using all data versus the subset of data for each time point (day 1, 3 and 5) for the BCR repertoire matrix. 157
Figure 32. Comparing co-expression networks derived using PCCs for day 1, day 3, day 5, all available data (ALLR) and an average of day 1 and day 5 PCCs (AVGR). 158
Figure 33. Determining the optimal linkage method for the BCR repertoire distance matrix using agglomerative coefficient (AC) score. 159
Figure 34. Determining the optimal number of clusters for the BCR repertoire distance matrix using fixed height cutting. 161
Figure 35. Dynamic Hybrid Tree Cutting. 163
Figure 36. rWGCNA identified 31 BCR repertoire modules. 165

Figure 37. Proportion of cluster variance explained by each BCR eigenvector (module).....166
Figure 38. PC1 feature loading scores for the 31 BCR repertoire modules. 168
Figure 39. Agglomerative hierarchical clustering of the 849 non-zero components from 10 iterations of SDA, applied to the imputed BCR repertoire matrix. 171
Figure 40. Feature loading scores and Posterior Inclusion Probabilities for the 41 robust BCR repertoire SDA components identified using hierarchical clustering. 172
Figure 41. The correlation between BCR modules (rWGCNA) and BCR components (SDA) calculated using repeated measures correlation (RMCORR). 174
Figure 42. The number (and percentage) of overlapping BCR features between strongly positively correlated SDA components (C) and rWGCNA modules (M). 176
Figure 43. AIC model selection: distribution of AICc scores for each BCR module.....201
Figure 44. Linear mixed effects models comparing the effect of disease status (health or sepsis) and time on BCR module score. 203
Figure 45. Partial residual plots showing the change in BCR module score in sepsis compared to health, across time, while controlling for covariates. 204
Figure 46. Stratification of sepsis patients using 7-day and 6-month mortality. 205
Figure 47. Linear mixed effect models investigating the effect of time and sepsis mortality group on BCR module score. 207
Figure 48. Partial residual plots showing the changes in BCR module score over time between sepsis mortality groups, controlling for covariates. 208
Figure 49. BCR module interaction network. 209
Figure 50. The effect of clinical variables on BCR module score in sepsis patients. 211
Figure 51. Linear mixed effects models exploring the effect of SRSq score on BCR module score (TSqASC). 213
Figure 52. Partial residual plots showing the dynamic changes in CIBERSORTx estimated cell fraction in health and sepsis, controlling for covariates. 215
Figure 53. Linear mixed effects models comparing the effect of disease status and time on CIBERSORTx estimated cell fraction. 216
Figure 54. Association between CIBERSORTx estimated cell fractions and BCR module score using cell fraction as the independent variable and module score as the dependent variable. 218
Figure 55. Partial residual plots showing the dynamic changes in immunoglobulin isotype gene expression (bulk RNA sequencing) during acute sepsis, controlling for covariates. 220
Figure 56. Linear mixed effects models investigating the effect of disease status and time on immunoglobulin isotype gene expression (bulk RNA-seq). 221
Figure 57. BCR module 1 summary. 223
Figure 58. CSR networks investigating the effect of time on CSR frequency in health and sepsis mortality groups. 226
Figure 59. CSR networks comparing sepsis mortality groups to health. 227
Figure 60. Quantitative trait analysis identified 451 genes significantly associated with BCR module 1 score. 230
Figure 61. Module 1 associated pathways. 233
Figure 62. Module 3 is elevated in sepsis and increases significantly during disease course. 235
Figure 63. SHM in non-class switched and class-switched isotypes in sepsis. 237
Figure 64. The distribution of SHM for IGHM and IGHD BCR sequences reveals two peaks in mutational load. 240
Figure 65. IgM^+ and IgD^+ BCRs were grouped into naïve and plasma or memory cell populations, according to the number of SHMs. 243
Figure 66. Module 3 associated pathways. 245
Figure 67. Comparing IGHV gene usage determined using either bulk RNA sequencing gene expression or targeted multiplex PCR amplification of the repertoire. 247
Figure 68. Quantitative trait analysis identified 47 genes significantly associated with BCR module 31 score. 250

Figure 69. Partial residual plots depicting biased IGHV5 gene usage associated with sepsis mortality, controlling for covariates. 253
Figure 70. Patient stratification by hierarchical clustering of the imputed BCR repertoire feature matrix. 254
Figure 71. BCR Principal Component Analysis. 255
Figure 72. Separation between health and disease is driven by a subset of BCR repertoire features. 255
Figure 73. Longitudinal B cell network from a patient with sepsis (LHS) and a healthy volunteer (RHS). 258
Figure 74. BCR clone network-level centralisation measures grouped by mortality for (A, B, E, F) degree centralisation and (C, D, G, H) eigenvector centralisation. 261
Figure 75. BCR degree centrality....262
Figure 76. Characterising persistent and central BCRs. 264
Figure 77. The Jaccard index was used to calculate the clonal overlap between the BCR repertoires at different time points. 265
Figure 78. Investigating B cell clonal dynamics in sepsis. 267
Figure 79. Longitudinal B cell clone isotype usage. 268
Figure 80. Longitudinal B cell clone isotype count. 269
Figure 81. QC summary: TCRα, β, γ and δ repertoire sequencing depth. 284
Figure 82. V gene detection in the TCR α (A), β (B), γ (C) and δ (D) repertoire datasets. 286
Figure 83. Agarose gel visualising the PCR product from TCRγ and TCRδ multiplex PCR amplification for the same 15 samples comprising libraries 10 and 11. 290
Figure 84. QC summary: TCR constant region usage for the technical replicates (healthy donor JR-1975-1003), grouped by sequencing lane. 291
Figure 85. QC summary: the percentage of total unique VDJ TCR reads for each V and J gene for the technical replicates (healthy donor JR-1975-1003), grouped by sequencing lane. 293
Figure 86. TRAV gene normalisation to account for batch effect detected in the technical replicates. 295
Figure 87. TRDV gene normalisation to account for batch effect seen in the technical replicates. 296
Figure 88. QC summary: TCR sequence overlap. 298
Figure 89. TCR rarefaction analysis. 299
Figure 90. Optimal statistical model determined by AIC model selection....306
Figure 91. Linear mixed effect models investigating the effect of time and disease status on TCRAB module score. 308
Figure 92. Partial residual plots showing the change in TCRAB module score in sepsis compared to health, across time, while controlling for covariates. 309
Figure 93. Linear mixed effect models investigating the effect of time and disease status on TCRGD module score. 311
Figure 94. Partial residual plots showing the change in TCRGD module score in sepsis compared to health, across time, while controlling for covariates. 312
Figure 95. Linear mixed effect models investigating the effect of time and sepsis mortality group on TCRAB module score. 314
Figure 96. Partial residual plots showing the changes in TCRAB module score over time, in sepsis mortality groups, while controlling for covariates. 315
Figure 97. Linear mixed effect models investigating the effect of time and sepsis mortality group on TCRGD module score. 316
Figure 98. Partial residual plots showing the changes in TCRGD module score over time, in sepsis mortality groups, while controlling for covariates. 317
Figure 99. TCRAB module interaction network. 318
Figure 100. TCRGD module interaction network. 319

Figure 101. The effect of A) clinical variables measured once per ICU admission on day 1 TCRAB module score and B) longitudinal clinical variables on TCRAB module score (all time points). ...321
Figure 102. The effect of A) clinical variables measured once per ICU admission on day 1 TCRGD module score and B) longitudinal clinical variables on TCRGD module score (all time points). ...323
Figure 103. Linear mixed effects models exploring the effect of SRSq score on TCRAB module score (TSqASC). ...325
Figure 104. Association between CIBERSORTx estimated cell fractions and TCRAB module score. ...327
Figure 105. Association between TCRGD module score and CIBERSORTx estimated cell fractions. ...328
Figure 106. TCRAB principal component analysis. ...329
Figure 107. Separation between health and disease is driven by a subset of TCRAB repertoire features. ...330
Figure 108. TCRGD principal component analysis. ...331
Figure 109. Separation between health and disease is driven by a subset of TCRGD repertoire features. ...332
Figure 110. Partial residual plots showing the significant increase in TCRAB module 3 score in sepsis compared to health, and the corresponding changes in TRBV gene usage for the constituent features. ...335
Figure 111. Genes with a significant effect on TCRAB module 3 score, from matched bulk RNA sequencing gene expression data. ...338
Figure 112. Partial residual plots showing the significant increase in TCRAB module 5 score in sepsis compared to health, corresponding to changes in T cell clonal expansion and diversification. ...341
Figure 113. Pathways associated with TCRAB module 1 score, identified using XGR. ...343
Figure 114. Partial residual plots showing the significant decrease in TCRAB module 9 score in sepsis early-death non-survivors compared to survivors and late-death non-survivors. ...345
Figure 115. Module 9 quantitative trait analysis. ...348
Figure 116. Partial residual plots showing the significant increase in TCRAB module 5 score in sepsis compared to health, corresponding to changes in γδ T cell subsets. ...350
Figure 117. Correlation between TCR and BCR modules, estimated using RMCORR. ...355
Figure 118. The number of VDJ IgH, TCR α, β, γ and δ sequences identified by TRUST4 for the ~900 sepsis samples for which bulk RNA sequencing data is available. ...376

List of Tables

Table 1. Sepsis-3 diagnostic criteria. 21
Table 2. Properties of immunoglobulin isotypes. 41
Table 3. Hypothesised mechanisms of BCR activation in response to antigen binding. 51
Table 4. Gamma-delta T cell subsets. 58
Table 5. BCR and TCR loci. 66
Table 6. The number of functional and non-functional V, D and J gene segments for the TCR and BCR. 69
Table 7. The Genomic Advances in Sepsis study inclusion criteria. 81
Table 8. The total number of V gene primers for BCR and TCR multiplex PCR reactions. 88
Table 9. TCR repertoire PCR protocol. 91
Table 10. Optimisation of the BCR and TCR multiplex PCR amplification protocol conducted as part of this thesis. 92
Table 11. BCR immunoglobulin isotype and TCR chain specific subsample depths. 105
Table 12. Clonal expansion and diversification indices. 106
Table 13. Statistical measures used to characterise the distribution of continuous repertoire traits. 107
Table 14. The classification of BCR sequences according to the total number of mutations. 109
Table 15. Classification of TCR sequences according to invariant TCRα V and J gene usage 111
Table 16. Classification of B cell clonal dynamics using the change in percentage of the repertoire between the first and last available time points. 120
Table 17. A compilation of commonly used summary measures for characterising adaptive immune repertoires. 126
Table 18. Specialised computational tools for analysis of adaptive immune repertoire sequencing data. 127
Table 19. QC summary statistics: BCR repertoire sequencing depth post-filtering, for all samples. 132
Table 20. Weighted gene correlation network analysis (WGCNA) summary. 146
Table 21. The three categories of missing data. 181
Table 22. BCR rWGCNA module summary table. 197
Table 23. Linear mixed effect models (LME) used in AIC model selection, simple effects tests (SE), mortality analysis (M) and SRSq (S) analysis. 199
Table 24. Predicted origin of IgM and IgD BCRs in health and sepsis, based on SHM burden and disease dynamics. 244
Table 25. Summary of node and graph-level network centrality measures. 257
Table 26. Sequencing timeline for the TCR repertoire libraries generated using multiplex PCR. 282
Table 27. QC summary statistics: TCRα, β, γ and δ repertoire sequencing depth. 285
Table 28. A summary of rWGCNA TCR specific optimisations. 301
Table 29. TCRAB module summaries annotated according to TCR α or β chain. 304
Table 30. TCRGD module summaries annotated according to TCR γ or δ chain. 304
Table 31. Significant perturbations in TRA and TRB V gene usage in sepsis. 334
Table 32. Genes associated with TCRAB modules 1, 4 and 5 which capture clonal expansion. 343

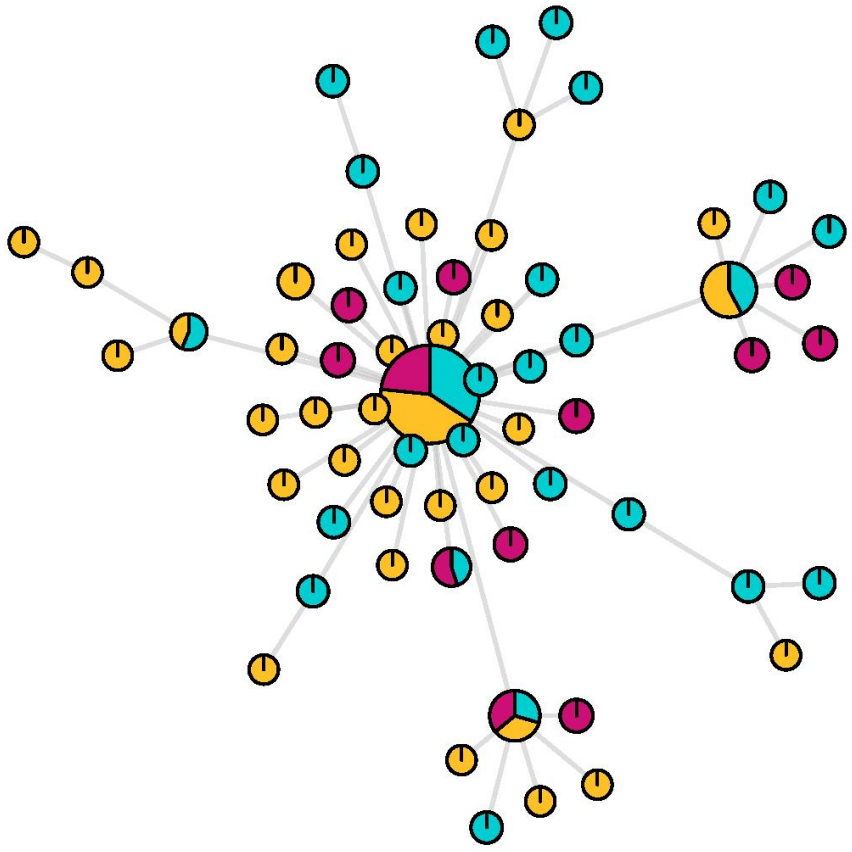
List of Equations

Equation 1. The Jaccard index was used to quantify V(D)J sequence sharing between samples. 103
Equation 2. Isotype usage estimated using targeted BCR repertoire sequencing. 111
Equation 3. A comparable measure of isotype usage was calculated using the TMM normalised isotype gene counts from bulk RNA sequencing. 112
Equation 4. Clonal overlap measured using the Jaccard index. 121

Nomenclature

AIC	Akaike Information Criterion
AIR	Adaptive Immune Repertoire
APACHE	Acute Physiology And Chronic Health Evaluation
BCR	B Cell Receptor
BH	Benjamini-Hochberg
CAP	Community-Acquired Pneumonia
CDR	Complementarity Determining Regions
CSR	Class Switch Recombination
DAMPS	Damage-Associated Molecular Patterns
eCRF	Electronic Case Report Form
ED	Early-Death
FDR	False Discovery Rate
FP	Faecal Peritonitis
FWR	Framework Region
GAinS	Genomic Advances In Sepsis
GC	Germinal Centre
GCS	Glasgow Coma Scale
ICU	Intensive Care Unit
IMGT	International Immunogenetics Information System
ITAM	Immunoreceptor Tyrosine Based-Activation Motif
JI	Jaccard Index
LD	Late-Death
LME	Linear Mixed Effect
MAR	Missing At Random
MRDARCY	MRD Assessment And Retrieval Code In Python
MST	Minimum Spanning Tree
NGS	Next Generation Sequencing
NSD	Normalised Standard Deviation
ORF	Open Read Frame
PAMPS	Pathogen-Associated Molecular Patterns
PBMC	Peripheral Blood Mononuclear Cells
PC	Principal Component
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PCR	Polymerase Chain Reaction
PRRs	Pattern Recognition Receptors
RMCORR	Repeated Measures Correlation
rWGCNA	Repertoire Weighted Gene Correlation Network Analysis
SAPS	Simplified Acute Physiology Score
SD	Standard Deviation
SDA	Sparse Decomposition Of Arrays
SHM	Somatic Hypermutation
SOFA	Sequential Organ Failure Assessment
SRS	Sepsis Response Signature
TCR	T Cell Receptor
TMM	Trimmed Mean Of M-Values
tSNE	T-Distributed Stochastic Neighbour Embedding
UMAP	Uniform Manifold Approximation And Projection For Dimension Reduction
UMI	Unique Molecular Identifier
VIF	Variance Inflation Factors
WCC	White Cell Count
WGCNA	Weighted Gene Correlation Network Analysis

Chapter 1. Introduction



1 Introduction

Chapter one presents the aims of this thesis and reviews the relevant existing scientific literature to establish the rationale for the proposed work and current state of knowledge in this area. Particular attention will be paid to summarising the adaptive immune response in sepsis, and the architecture of the adaptive immune repertoire including the generation of receptor diversity via V(D)J recombination. B and T cell receptor repertoire sequencing methods will also be discussed.

1 Introduction 17

1.1 Sepsis 18

1.2 The Adaptive Immune System 32

1.3 The Adaptive Immune Repertoire 66

1.4 Chapter 1 Aims 77

Sepsis is the dysregulated host response to infection, resulting in organ dysfunction and in many cases, death. Long-term immune dysregulation occurs following the acute episode, which is associated with increased susceptibility to secondary infection, implicating the adaptive immune system in disease progression and resolution. The overarching aim of this thesis is to characterise and understand the adaptive immune repertoire in sepsis caused by community-acquired pneumonia (CAP), during acute disease.

To achieve this aim, I have generated B and T cell receptor repertoire (BCR/TCR) data for a large cohort of sepsis patients, using an in-house multiplex PCR reaction combined with next generation sequencing (NGS). I also developed a novel computational approach for repertoire analysis based on weighted gene correlation network analysis (WGCNA) termed repertoire WGCNA (rWGCNA). I have applied this pipeline to the sepsis data, providing insight into disease pathogenesis. The research described in this thesis was conducted as part of the UK Genomic Advances in Sepsis study (GAinS, www.ukccg-gains.org).

To date, there has been limited research in this area, with existing studies primarily relying on outdated and low-throughput molecular methods. This introduction will provide an overview of the relevant literature pertaining to the objectives of this thesis. To begin, I present an introduction to sepsis, emphasising disease heterogeneity and discussing the importance of clinical microbiology in understanding pathophysiology. Next, I describe the structure and function of the adaptive immune system, focusing on B and T cell response to antigen. Additionally, I review the discovery of novel biomarkers and potential therapeutic strategies centred on the adaptive immune system in the context of sepsis. I then introduce the adaptive immune repertoire, focusing on the generation of repertoire diversity through V(D)J recombination and experimental methods for repertoire analysis. Finally, I outline the specific aims of my project.

1.1 Sepsis

1.1.1 Introduction and Disease Burden

Sepsis is defined as the life threatening, dysregulated host response to infection and is associated with organ dysfunction, high mortality and in survivors, long-term morbidity and mortality (Singer et al., 2016, Schroeder and Cavacini, 2010). The current clinical definition of sepsis captures a complex and heterogeneous patient cohort due to the interaction of host, pathogen and environmental factors [**Figure 1**], making diagnosis challenging, even for experienced clinicians. Mortality rates are high; estimates range from 30 to 50% (Sakr et al., 2018). More than 30 million people are affected by sepsis worldwide each year, resulting in upwards of 6 million deaths. However, this is likely an underestimate, the Global Burden of Disease study argues 50 million cases and 11 million deaths per year are more accurate, as data from low and middle-income countries is limited (Rudd et al., 2020, Wiersinga and van der Poll, 2022).

Sepsis represents a substantial economic burden; in the USA treatment costs on average $22,100 and in 2017 it was one of the most expensive causes of hospital inpatient care across health care payers (*e.g.* Medicare and Medicaid). In the United Kingdom (UK) sepsis is the leading cause of adult intensive care unit (ICU) admission and sepsis related care costs approximately £10 billion per year nationally (Bray et al., 2020, Liang et al., 2006). The incidence of sepsis is increasing and costs are expected to grow without advancements in targeted therapeutics. Rising case numbers can be attributed, in part, to improved reporting and diagnosis, but an ageing, vulnerable population with predisposing co-morbidities such as obesity and diabetes, drug-resistant pathogens and the increasing frequency of invasive procedures are also expected to contribute (Rhee and Klompas, 2020).

Sepsis is considered the final common pathway to death for most infectious diseases (Cecconi et al., 2018). However, sepsis-specific therapies are limited to organ supportive care and broad-spectrum antimicrobials; no targeted immunomodulatory therapies exist, highlighting an unmet clinical need (Angus et al., 2001, Vincent et al., 2006, Taeb et al., 2017). Consequently, in 2017, the World Health Organisation (WHO) declared sepsis a global health priority, aiming to improve prevention, diagnosis and disease management (Reinhart et al., 2017).

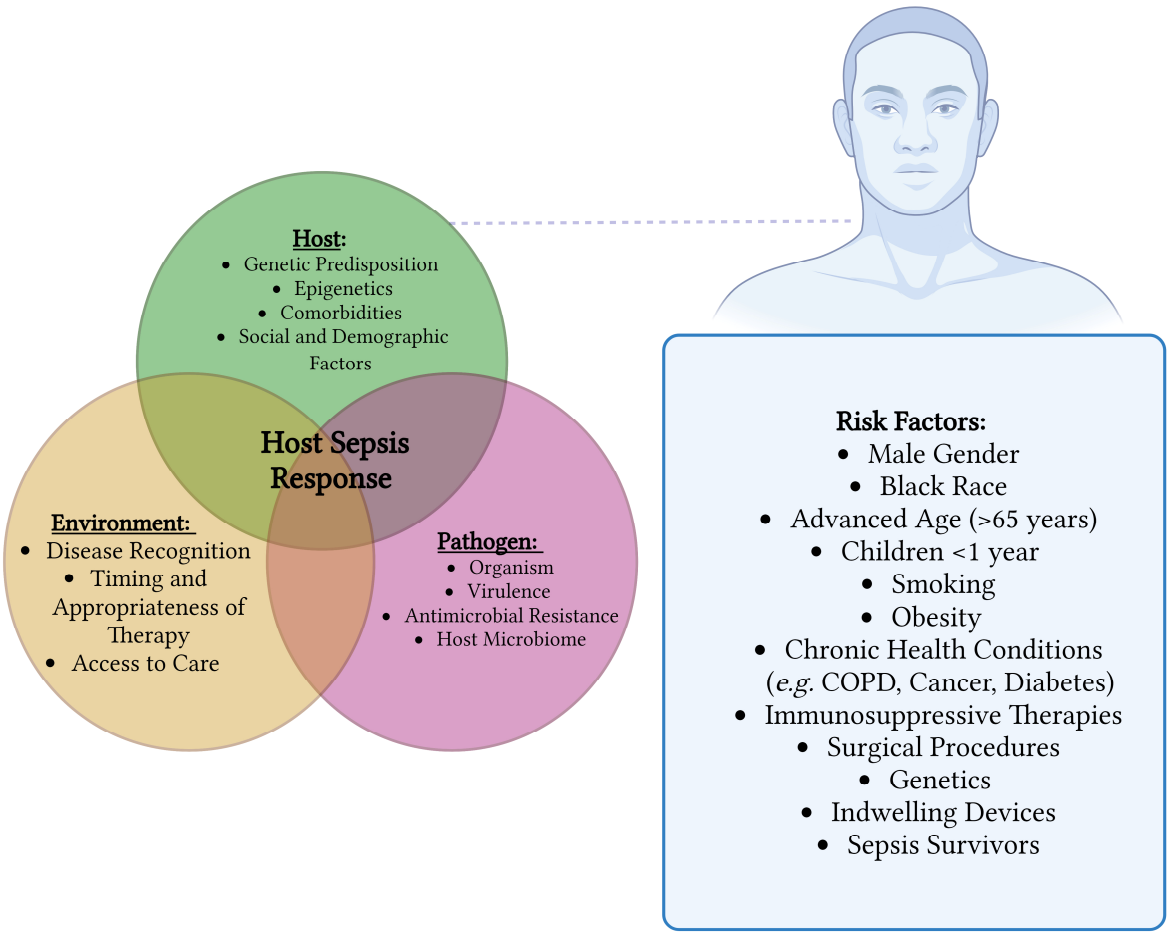


Figure 1. Heterogeneity in sepsis.

Sepsis disease heterogeneity is driven by the interaction between host, pathogen and environmental factors and can be observed at the molecular and/or physiological level, influencing treatment and outcome.

1.1.2 Clinical Definition

The definition of sepsis was last revised in 2016 with the publication of the Third International Consensus Definitions for Sepsis and Septic Shock. Prior to this, the definition of sepsis focused predominantly on the systemic inflammatory response under the prevailing model that sepsis follows a trajectory from severe sepsis to shock (Singer et al., 2016) [**Table 1**]. As anti-inflammatory drugs failed to improve survival rates and evidence of persistent immunosuppression was increasingly documented, a reinterpretation of the disease was prompted and the Sepsis-3 guidelines were published. Today, sepsis is considered a complex,

simultaneous dysregulation of pro and anti-inflammatory processes, occurring in response to infection and shaped by host and pathogen factors (Hotchkiss et al., 2013).

Organ dysfunction:	An increase of >2 points in the Sequential Organ Failure Assessment (SOFA) score or 'quick' qSOFA.
Hypotension:	Systolic blood pressure ≤100mgHg.
Altered mental status	Glasgow Coma Scale score of ≤13
Tachypnoea	Respiratory rate ≥22 beats/minute
Evidence of infection	*E.g.* respiratory symptoms or purulent wounds.
Septic shock	Circulatory and cellular/metabolic alterations profound enough to substantially increase mortality, persistent hypotension requiring vasopressors to maintain mean arterial pressure (MAP) ≥65 mmHg and serum lactate >2mmol/L despite resuscitation.

Table 1. Sepsis-3 diagnostic criteria.

As published in the 2016 Third International Consensus Definitions for Sepsis and Septic Shock (Singer et al., 2016).

1.1.2.1 Disease Trajectory and Post-Sepsis Syndrome

Historically, sepsis was considered a biphasic disease with two peaks in mortality [**Figure 2**]. The first phase occurred in the week post-onset and was thought to represent a period of overwhelming inflammation and cytokine storm, causing organ failure. The second phase was a period of immune suppression occurring 3-4 weeks later, predisposing the patient to secondary infection and viral reactivation (Denstaedt et al., 2018, Mathias et al., 2015, van der Slikke et al., 2020). 'Recovery' was marked by survival of these episodes. Nowadays, it is understood that pro and anti-inflammatory responses interact simultaneously in sepsis (**1.1.2**). Improvements in clinical care (notably early recognition and prompt antimicrobial therapy) have reduced 30-day mortality and revealed that 'recovered' individuals remain at an increased risk of death for many months after acute disease due to protracted immune suppression (Denstaedt et al., 2018, Cao et al., 2019). In the year post-sepsis, 60% of

survivors are re-hospitalised and 1 in 6 individuals die (Shankar-Hari and Rubenfeld, 2016). Prescott et al. (2016) report a 22.1% increase in mortality in sepsis-survivors relative to non-hospitalised controls, and a 10.4% and 16.3% increase in mortality relative to patients with non-septic infection and sterile inflammation respectively. Elevated mortality endured throughout the two-year follow-up period and was largely explained by hospital re-admission for non-sepsis infection or sterile inflammation. Rahmel et al. (2020) report similar findings in a five-year retrospective matched cohort study. Specifically, five-year mortality after sepsis and to a lesser extent severe infection, was elevated compared to matched controls.

'Post-sepsis syndrome' is the term used to describe the long-term physical and psychological effects of sepsis and is an area of limited research. Notably, the excess mortality reported by Rahmel et al. (2020) was significantly reduced for patients receiving after-care treatment in a rehabilitation facility, highlighting the importance of understanding post-sepsis syndrome. Features of immune senescence in response to normal physiological ageing include impaired lymphopoiesis and reduced B cell diversity, which are similarly observed in sepsis survivors (Shankar-Hari and Rubenfeld, 2016). This suggests that long-term immunological impairment in sepsis may reflect a state of accelerated immune-senescence.

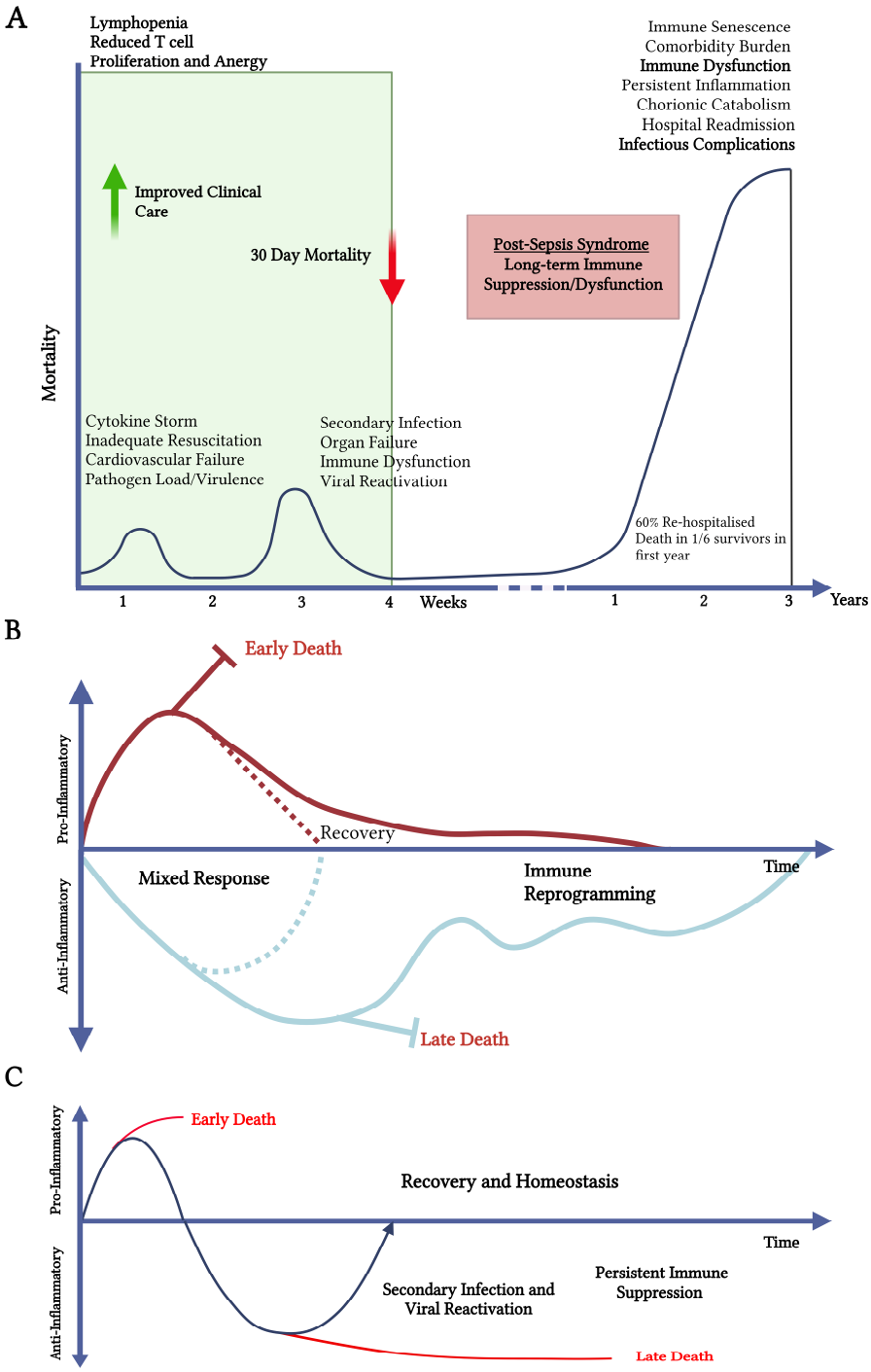


Figure 2. Sepsis disease trajectory.

A) Mortality in sepsis: advances in clinical care and increased survival rates have revealed a third peak in mortality, occurring months after the initial sepsis episode. This peak is associated with persistent immune suppression and susceptibility to secondary infection. **B)** The revised sepsis disease model (Third International Consensus Definitions for Sepsis and Septic Shock): according to this updated model, sepsis involves the complex interaction between pro and anti-inflammatory processes. This is superseded by a period of protracted immune suppression. Although early peaks in mortality exist, they are reduced compared to historical rates. **C)** The historical model of sepsis: in the past sepsis was characterised by two distinct phases, the pro-inflammatory response followed by a period of immune suppression. Figure adapted from van der Slikke et al. (2020), Denstaedt et al. (2018), Mathias et al. (2015), Delano and Ward (2016) and Hotchkiss and Karl (2003).

1.1.3 Dysregulation of the Inflammatory Response in Sepsis

Human immune response to infection is initiated by the presence of pathogen-associated molecular patterns (PAMPs) such as endo and exo-toxins, lipopolysaccharide (LPS), viral double and single stranded-RNA, microbial DNA and host derived damage-associated molecular patterns (DAMPs) which are released during stress *e.g.* mitochondria and heat shock proteins released from injured tissue (Wiersinga et al., 2014, Wiersinga and van der Poll, 2022). These molecules bind to pattern recognition receptors (PRRs) including Toll-like receptors (TLR), Nod-like receptors (NLRs) and C-type lectin receptors (CLR) expressed on antigen-presenting cells and monocytes as well as epithelial barriers. Receptor recognition promotes a cascade of activation which triggers translocation of nuclear factor–ϰB (NF-ϰB) and initiation of an intracellular transduction pathway which regulates the transcription and release of immunoregulatory effector cytokines such as pro-inflammatory tumour necrosis factor (TNF)-α, interleukin (IL)-1, IL-6, IL-12 and interferon (IFN)-γ. It also promotes the secretion of nitric oxide (NO) and the activation and recruitment of other immune cells which initiates innate immunity and causes a localised immune response (Jarczak et al., 2021, Gyawali et al., 2019). Pro and anti-inflammatory pathways are activated simultaneously, promoting the secretion of multiple cytokines and chemokines.

In a normal immune response to infection, equilibrium between pro and anti-inflammatory processes is reached, maximising pathogen clearance and enabling the repair of damaged tissue. However in sepsis, equilibrium is not maintained; a localised infection becomes systemic, PRRs are persistently stimulated and both anti and pro-inflammatory processes themselves become detrimental, variable and unpredictable. This process is termed a 'cytokine' storm (Schulte et al., 2013). The dysregulated immune response leads to widespread physiological effects across organs and tissue types (Lonsdale et al., 2020).

The complement cascade is activated, enhancing the inflammatory response while components of the adaptive immune system are downregulated (van der Poll et al., 2017). Cumulative inflammation causes progressively greater tissue damage and organ dysfunction; responsible for much of the morbidity and mortality in sepsis. The complex interaction of pro-inflammatory and coagulation cascades causes a range of pathology, from mild thrombocytopenia to fulminant disseminated intravascular coagulation (DIC). Clotting is exaggerated by rapid release of neutrophil extracellular traps (NETs) for pathogen capture, which inadvertently cause platelet aggregation, thrombosis and contribute to ongoing inflammation and cellular exhaustion. While platelets are known to maintain vascular integrity in a controlled immune response, in sepsis excessive platelet activation is associated with organ dysfunction. Disruption to the endothelial barrier is common due to cytokine release; intravascular fluid leaks into the interstitial space causing hypo-perfusion of tissues and organs which presents as the characteristic skin mottling and increases accessibility for pathogens and their products to invade the host (Schulte et al., 2013). Hypo-perfusion promotes a shift to anaerobic glycolysis, which produces lactic acid and is associated with significant morbidity and mortality (Gyawali et al., 2019, Flynn Makic, 2019).

1.1.4 Sepsis Immunosuppression

Sepsis-related hyper-inflammation is superseded by immune-suppression; deficits occur in both innate and adaptive immunity, preventing pathogen clearance. Autopsies have revealed that many deceased individuals have unresolved opportunistic infections. A predominant feature of sepsis immunosuppression is global depletion of lymphocytes due to apoptosis of $CD4^+$ and $CD8^+$ T cells, B cells and follicular dendritic cells, T cell exhaustion (due to chronic high-load antigen exposure) and expansion of regulatory immunosuppressive T_{reg} cells and myeloid derived suppressor cells (van der Poll et al., 2017, Stearns-Kurosawa

et al., 2011, Hotchkiss et al., 2013). Neutrophils exhibit a decreased ability to clear bacteria and be recruited to the site of infection and demonstrate delayed apoptosis. The presence of immature neutrophils, with deficits in microbial effector functions has also been observed and these changes persist into convalescence (Kwok et al., 2023). Antigen-presenting cells have reduced HLA-DR expression, an impaired ability to secrete pro-inflammatory cytokines and are vulnerable to apoptosis (van der Poll et al., 2017). Epigenetic regulation of gene expression is an important mechanism for establishing immune suppression in late sepsis and is characterised by changes in histone methylation and acetylation and loss of activation marks in promotors (van der Poll et al., 2017, Falcao-Holanda et al., 2021). Dysregulation of the adaptive immune system in sepsis will be discussed in more detail in section **1.2.1**.

1.1.5 Sepsis Aetiology and Epidemiology

Consensus opinion is that sepsis pathophysiology occurs via one or more common pathways regardless of infectious agent, although the pathogen can influence treatment and outcome. Circulating infectious diseases vary across geographical area and patient demographics. As estimates suggest 80% of hospital-treated sepsis cases arise in the community, infectious aetiology significantly contributes to disease heterogeneity, posing a clinical challenge (Michaud, 2009). Worldwide, lower respiratory infections are a 'top ten' cause of death and this is reflected in sepsis. The leading site of infection for sepsis-associated mortality was the lung from 1990-2017 (Rudd et al., 2020, Michaud, 2009). Intra-abdominal sites, the urinary tract and soft tissue also comprise a substantial proportion of infection sites (Abe et al., 2019, Karlsson et al., 2007, Rudd et al., 2020, Tridente et al., 2014).

According to Sepsis-3 guidelines [**Table 1**], a diagnosis of sepsis requires evidence of infection, but only 40-60% of blood cultures are positive following hospital admission (Singer

et al., 2016). Where a pathogenic organism is identified, the predominant cause of sepsis is bacterial. The Sepsis Occurrence in Acutely Ill Patients (SOAP) study reported an equal prevalence of gram-positive and gram-negative bacterial infections in sepsis. These were frequently *Staphylococcus aureus, Pseudomonas* or *Escherichia coli* (Dolin et al., 2019, Vincent et al., 2006, Tridente et al., 2014). However, estimates in the literature vary; the SOAP study linked fungal infections to ~10% of sepsis cases, while others estimate the case-burden to be between 4-20% (Karlsson et al., 2007, Vincent et al., 2009). In Finland, gram-positive infections are more common but the opposite was reported in the USA by the 2012 Intensive Care over Nations Study. Differences in recruitment strategy and location may explain discrepancies (Vincent et al., 2009, Karlsson et al., 2007, Sakr et al., 2018). In Europe, the leading infectious cause of ICU admission is community-acquired pneumonia (CAP) often from *Streptococcus pneumoniae* infection. The failure to detect an etiological agent in more than a third of sepsis cases is found across multiple studies and impedes our ability to understand the role of pathogen heterogeneity in relation to disease heterogeneity (Walden et al., 2014). In the clinic this limits the opportunities to provide targeted treatment.

Similarly, almost any virus can cause sepsis in susceptible individuals. This may explain a significant proportion of blood culture negative cases. The reported incidence of viral sepsis is low, but viruses are not routinely tested for which may explain the inconsistency (Vincent et al., 2006). Herpes Simplex Viruses (HSV) and enteroviruses are the most common viruses identified in neonatal sepsis and influenza viruses frequently cause sepsis in older adults, immunocompromised individuals and pregnant women (Shane et al., 2017, Lin et al., 2018). Although bacteria are the most frequent cause of CAP sepsis, in the CAP-China study more than a third of patients admitted to hospital with either influenza or non-influenza viral

infections went on to develop sepsis, suggesting routine viral testing would be beneficial in the clinical setting (Gu et al., 2020).

Several factors increase the likelihood of developing sepsis. However, well understood risk-factors for sepsis are predominantly those that predispose someone to developing infection including age, male sex, black race and chronic health conditions or use of immunosuppressive therapies [**Figure 1**]. More than half of severe cases arise in individuals over 65 years old and more than half of patients have at least one other health condition such as chronic obstructive pulmonary disease (COPD), cancer or diabetes (Mayr et al., 2014). Genetic factors that predispose to the development of the severe, dysregulated immune response and inability to clear infection, which are characteristic of sepsis, are less understood.

Research has shown that susceptibility to infection is highly heritable, for example adult adoptees have a five-fold higher risk of dying from infection if a biological parent died of infection before the age of 50 years, indicating the presence of genetic variants which predispose to severe infection (Engoren et al., 2022). Exome sequencing of blood samples from paediatric patients with severe sepsis and no prior family history of immunodeficiency revealed the presence of novel pathogenic variants in primary immunodeficiency genes in 2/8 fatal cases, implicating rare genetic variants in sepsis (Asgari et al., 2016). Additionally, Orange (2005) identified several primary immunodeficiency disorders associated with increased risk of developing sepsis including Hyper-IgM syndrome and common variable immunodeficiency (CVID) which is characterised by low levels of protective antibodies. Genome wide association studies (GWAS) have been used to identify genetic risk factors for sepsis, but due to disease heterogeneity, have had limited success. To date, only a small number of genes have been identified including genetic polymorphism rs4957796 within the

FER gene, which encodes a non-transmembrane receptor tyrosine kinase (Rautanen et al., 2015). TT homozygosity is associated with elevated 90-day mortality in CAP sepsis (Hinz et al., 2017). Improvements in patient stratification and mass recruitment have enabled more complex analysis (Rosier et al., 2021, Guillen-Guio et al., 2020). Engoren et al. (2022) genotyped >2000 sepsis patients and ~13,000 controls and developed polygenic risk scores able to predict sepsis with good discrimination. They also identified variants independently associated with sepsis patients recruited under current Sepsis-3 guidelines versus Sepsis-2 guidelines, which suggests a genetic basis to disease heterogeneity. However, the authors note that sample size was insufficient to address hypotheses linking genetic and clinical heterogeneity.

1.1.6 Clinical Microbiology

Rapid identification of the infectious agent in sepsis directs treatment. Microbiology techniques including cell culture, gram staining and biochemical tests are the 'gold standard' and many are automated (Chun et al., 2015). In sepsis, blood cultures are the primary diagnostic method and as discussed in section **1.1.5**, the culture negative rate approaches 50% due to a variety of factors including undiagnosed viral infection, the use of antibiotics in the preceding 48 hours and challenges culturing specific bacterial and fungal species (Sigakis et al., 2019). New molecular techniques, including PCR-based tests and nucleic acid sequencing will provide more accurate diagnoses at higher sensitivity; beneficial to researchers interested in understanding disease heterogeneity.

1.1.6.1 Multiplex PCR

Multiplex PCR enables the simultaneous detection of multiple pathogens including bacteria and fungi, predetermined by a panel of primers targeting conserved regions of the genome

(Westh et al., 2009, Trung et al., 2019). Commercial kits are available; the MagicPlex Sepsis Real Time Test (Seegen, Seoul, Korea) for whole blood samples has a specificity of over 90% for 90 different pathogens and can also screen for drug resistance markers. However it has poor sensitivity (29%) so cannot be used independently of other tests (Zboromyrska et al., 2019). The Hyplex Blood screen system (BAG, Lich, Germany) uses multiplex PCR amplification with biotinylated primers to target pathogen specific housekeeping genes from 10 species and achieves nearly 100% sensitivity. However, high sensitivity comes at the cost of a lengthy multi-step protocol with high amplicon carry-over contamination risk (Wellinghausen et al., 2004). The disadvantage of PCR based tests is their reliance on pre-specified primer panels which are limited in their scope. Panels may fail to consider epidemiology or intra-species variation resulting in false-negative results.

1.1.6.2 Clinical Metagenomics

Clinical metagenomics is the characterisation of all DNA or RNA present in a sample using next generation sequencing, enabling simultaneous analysis of host and pathogen genomes or transcriptomes (Chiu and Miller, 2019). Decreasing costs, improved analysis tools and increasingly well-curated pathogen databases have established its use in clinical samples. In brief, sequencing reads are mapped to both the human genome and also a broad reference of pathogen genomes for unbiased, comprehensive identification and quantification of the organisms within a sample including bacteria, fungi, parasites and viruses. As non-human reads usually comprise less than 1% of the total sequencing reads, sensitivity can be a critical issue. Although NGS typically yields more positive organism calls than blood cultures, interpreting complex metagenomic sequencing data in a way that is clinically meaningful can be challenging. Discerning causative versus non-pathogenic/commensal organisms and differentiating between active and latent infection at low read depth is difficult, limiting

diagnostic potential and requiring input from a multidisciplinary precision medicine team (Duan et al., 2021, Long et al., 2016, Ren et al., 2021). False negatives due to intra-species variation are a common problem.

An alternative NGS approach is the differentiation of bacterial and viral sepsis based on the expression of host genes involved in bacterial or viral response pathways. Sweeney et al. (2016) developed a gene expression based 'Sepsis MetaScore', which when paired with a classifier using annotated bacterial versus viral induced genes, ascertained whether antibiotics should be administered, although clinical validation was not performed.

1.1.7 Genomic Advances in Sepsis Study

1.1.7.1 Introduction

In 2005, the UK Critical Care group established the Genomic Advances in Sepsis study (GAinS) (co-chief investigators: Professors Hinds and Knight, www.ukccggains.com) to investigate the genetic underpinnings of susceptibility to and outcome in sepsis. By the time recruitment ended in 2018, more than 2000 patients had been enrolled from 34 ICUs at 27 locations across the UK. Recruitment primarily focused on individuals with severe sepsis from community-acquired pneumonia (CAP) or faecal peritonitis (FP).

1.1.7.2 Sepsis Response Signatures

An important finding from the GAinS study was the discovery of two sepsis patient subgroups, identified by unsupervised hierarchical clustering of whole blood leukocyte gene expression micro-array data. Subgroups were termed Sepsis Response Signature 1 and 2 (SRS1/2) and can be determined by differential expression of 7 genes (*DYRK2, CCNB1IP1, TDRD9, ZAP70, ARL14EP, MDC1* and *ADGRE3*) (Davenport et al., 2016). More recently, a quantitative SRS score (SRSq), bounded between 0 and 1, with lower values indicating a

patient is transcriptionally closer to health and higher values indicating similarity to the highest extreme of SRS1 has been developed (Cano-Gamez et al., 2022).

SRS1 is associated with elevated mortality, indicators of severe disease and immune suppression including endotoxin tolerance (defined as the profoundly altered cellular response to endotoxin after repeated exposure), T cell exhaustion and metabolic derangement (West and Heagy, 2002). The SRS2 group is comparatively immunocompetent, and represents 60% of patients. Evidence suggests SRS groups are clinically relevant, for example differential response to corticosteroid treatment has been observed between groups. SRS2 patients have significantly higher mortality when treated with corticosteroids, while no difference is seen in SRS1 patients (Antcliffe et al., 2019). Kwok et al. (2023) report that immature and immunosuppressive neutrophils are the likely drivers of the SRS1 sub-phenotype, supporting the definition of SRS1 as an immunocompromised disease endotype.

Both SRS and SRSq scores have applicability beyond sepsis, having been used to stratify patients into clinically relevant subgroups in multiple different infectious diseases, including influenza, acute respiratory distress syndrome, COVID-19 and in paediatric septic shock. Hence, ongoing work aims to develop a rapid clinical assay for SRS assignment using qPCR (Cano-Gamez et al., 2022).

1.2 The Adaptive Immune System

1.2.1 Pathology in Sepsis

The adaptive immune system is comprised of multiple cell types including B cells, T cells and natural killer T cells which regulate the immune response. Impairments in adaptive immunity increase the likelihood of developing persistent, recurrent, secondary and nosocomial infections. The adaptive immune system responds slowly to antigen exposure

compared to the innate system, but can recognise and mount a highly specific response to an almost limitless range of antigens. In sepsis, the adaptive immune system is markedly dysregulated [**Figure 3**].

Apoptosis is a hallmark of sepsis and causes a major depletion of adaptive immune cells including $CD4^+$, $CD8^+$ and γδ T cells and B cells across all age groups irrespective of infectious agent. The mechanisms underpinning lymphopenia are poorly understood (Brady et al., 2020, Boomer et al., 2012). One theory proposes that sepsis disrupts the apoptosis-induced decrease in potent dendritic cells resulting in impairments to innate and adaptive immunity. Alternatively, factors released during sepsis including steroids and cytokines (*e.g.* TNF-α) dysregulate apoptosis via caspase-8 activity in the death-induced signalling complex or modulate survival factors in the Fas apoptotic or mitochondrial pathways (Cao et al., 2019). Research indicates that multiple cell death stimuli are present in sepsis; pharmacological targeting of any one apoptotic trigger is unlikely to restore homeostasis (Hotchkiss et al., 2013). Uncontrolled apoptosis reduces immune surveillance and drives systemic immunological tolerance with anti-inflammatory properties, causing hypo-responsiveness and vulnerability to infection (Bomans et al., 2018). Whether dysregulation of apoptosis is a pathophysiological mechanism or side effect is debated (Hotchkiss et al., 2013). Surviving $CD4^+$ and γδ T cells show reduced cytokine secretion (IL-2, IL-12, IFN-γ, and CD69). $CD8^+$ T cells have decreased cytotoxic activity and express markers of exhaustion (Brady et al., 2020). The percentage of exhausted $CD21^+$ B cells is also elevated and selective depletion of memory B cells has been observed (Brady et al., 2020).

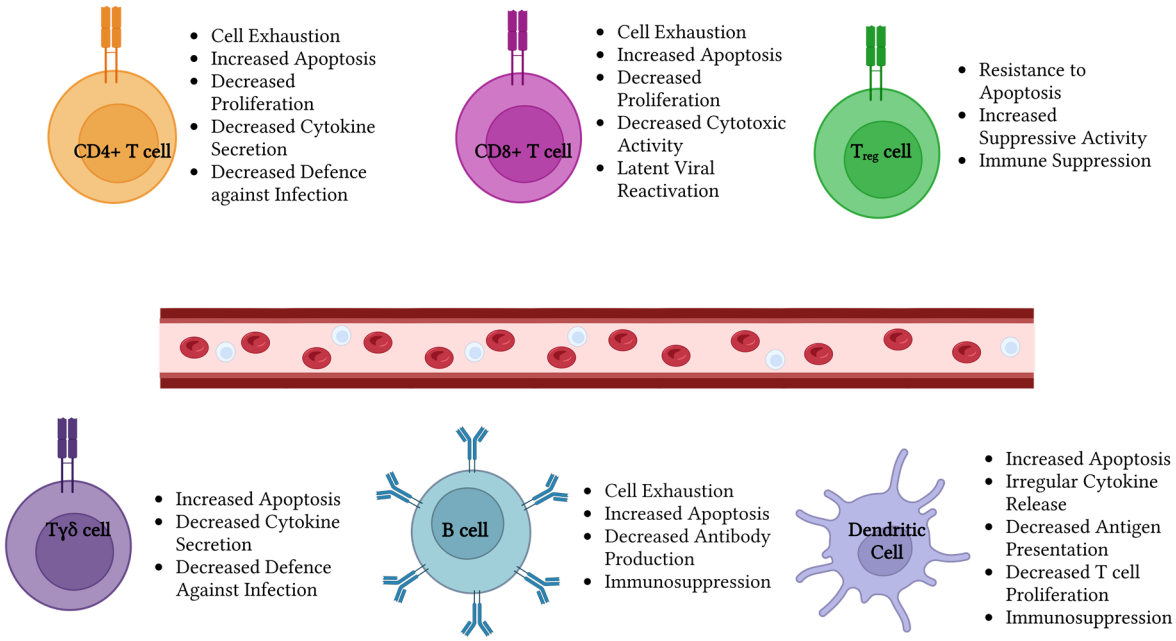


Figure 3. The dysregulated adaptive immune response in sepsis.

A universal characteristic of the dysregulated adaptive immune response in sepsis is increased apoptosis across the majority of cell types with the exception of T_{reg} cells. Immune suppression is evident and cells express markers of exhaustion. T_{reg} cells exhibit elevated expression of anti-apoptotic markers and are able to evade cell death. The relative increase in T_{reg} cell numbers negatively impacts T cell proliferation and further contributes to immune suppression. Figure adapted from Brady et al. (2020).

1.2.2 Biomarkers in Sepsis

A variety of biomarkers have been explored for their prognostic potential in sepsis including: cytokines, chemokines, DAMPs, non-coding RNAs and metabolites. In sepsis, greater B and T cell apoptosis is associated with worse outcome; a simple measure of $CD4^+$ T cell count at admission can identify patients at risk of in-hospital death (Polilli et al., 2021).

Several researchers have explored B and T-lymphocyte Attenuator (BTLA) as a dynamic, prognostic marker in sepsis. BTLA is an inhibitory molecule expressed on B and T cells, macrophages, dendritic and NK cells and plays an important role in immune regulation. BTLA binding to its ligand, Herpes virus entry mediator (HVEM) inhibits T cell proliferation and differentiation and reduces cytokine production. BTLA also exists in soluble form (sBTLA), which antagonises negative co-signalling by competitively binding to

HVEM. In sepsis, sBTLA levels greater than 21ng/µl are associated with a 5-fold increase in mortality and correlates with disease severity (Lange et al., 2017). As a biomarker, sBTLA is limited to disease severity as it cannot independently distinguish sepsis from a critically-ill cohort and the underlying biology is poorly understood (Monaghan et al., 2018, Lange et al., 2017).

Over the past ten years, immunoglobulins have been a focus for biomarker discovery in sepsis as reduced serum concentrations in sepsis are a recurring finding in the literature. Low levels of circulating IgG1, IgM and IgA correlate with increased mortality in severe sepsis, which has paved the way for use of targeted immunoglobulin-replacement therapy (Bermejo-Martin et al., 2014). Circulating IgM has been explored as a dynamic marker of disease progression as a reduction in concentration often coincides with progression to septic shock. Giamarellos-Bourboulis et al. (2013) hypothesise that in sepsis, B cells are hypo-functional for IgM production, and this is compensated for by circulating serum IgM. When septic shock develops, serum IgM is fully consumed and lymphocytes are anergic to IgM production; survival is perhaps linked to the ability to maintain an appropriate body IgM distribution (Bermejo-Martin and Giamarellos-Bourboulis, 2015). However, immunoglobulin serum concentrations have been explored as biomarkers across many different diseases, suggesting their use in sepsis, while informative, is non-specific. For example, Monroy-Iglesias et al. (2022) performed a systematic review of immunoglobulins and site-specific cancer risk and identified several associations with the serum concentrations of specific isotypes, although these were inconsistent across datasets. Interestingly, no significant associations were found for IgM.

1.2.3 Targeted Therapeutics in Sepsis

Despite many clinical trials there are no effective pharmacological treatments for sepsis. Instead, prompt recognition and treatment focused on eliminating the source of infection (antifungals, broad-spectrum antibiotics, drainage of infected fluids) is key. Every hour delay in effective antimicrobial treatment decreases survival by ~7% (Kumar et al., 2006). Organ supportive care, including intravenous fluids to manage hypotension, should commence immediately upon recognition and a measure of serum lactate can be used as a marker of disease severity. Vasopressors to maintain tissue perfusion, dialysis and ventilation are used when appropriate (Gyawali et al., 2019).

The combination of both pro and anti-inflammatory factors poses a challenge to drug development. Therapies targeting excessive inflammation must be administered within a narrow timeframe after disease onset, specifically to those with elevated pro-inflammatory cytokines. In developed healthcare systems, sepsis survivors are frequently readmitted to hospital with new infections and sepsis episodes (discussed in section **1.1.2.1**). However, as secondary infections are only attributed to 10.9% of sepsis mortality, targeted use of immune stimulants in those with immune-suppression rather than hyper-inflammation is required (van der Poll et al., 2017).

Historically, several therapies have attempted to neutralise the excessive inflammatory response, including monoclonal antibodies against TNF, IL-1 receptors and endotoxins however these have not consistently improved survival across patient cohorts (Ziegler et al., 1991, Fisher et al., 1996, Stearns-Kurosawa et al., 2011). Human recombinant activated protein C therapy was introduced to reduce the risk of death associated with severe sepsis and septic shock, as decreases in Protein C exacerbate blood clotting. A review of six clinical trials found no evidence of efficacy and an increased risk of bleeding, therefore clinical trials

were discontinued (Marti-Carvajal et al., 2012). Corticosteroids are commonly administered to modulate excessive inflammation but usage in sepsis confers little overall reduction in mortality (~2%) (Rochwerg et al., 2018). Corticosteroids may even increase mortality in patients with overwhelming infection or the SRS2 (relatively immune-competent) phenotype (Antcliffe et al., 2019).

Currently, most individuals survive the hyper-inflammatory phase and develop protracted immunosuppression. Therefore, adaptive immune system stimulants, which are a relatively new pharmacological approach, may benefit a wider range of patients. Pro-survival cytokines such as IL-7 and IL-15 are a safe and effective means of boosting lymphocyte counts (Boomer et al., 2014). Recombinant human IL-7 (CYT107) increased $CD4^+$ and $CD8^+$ T cell counts, proliferation and activation for weeks after drug administration but did not alter mortality in sepsis (Francois et al., 2018). Most patients up-regulate the PD-1/PD-L1 pathway during the transition from hyper to hypo-inflammation. The PD-1/PD-L1 pathway regulates T cell exhaustion and blocking this interaction prevents apoptosis. An anti-PD1 IgG4 monoclonal antibody was trialled in sepsis with promising results, restoring immune status over 28 days and preventing lymphocyte apoptosis (Hotchkiss et al., 2019).

1.2.4 B Cell Biology

1.2.4.1 Introduction

B cells play a pivotal role in regulating the immune response and eliminating pathogens by producing and secreting antibodies, acting as antigen presenting cells and regulating other cell types including T cells. Dysregulation of B cells may disrupt these essential functions contributing to the pathogenesis of autoimmune diseases (Carter, 2006). The antibody

mediated B cell response is initiated by a membrane-bound immunoglobulin called the B cell receptor (BCR). Each B cell expresses multiple copies (10^5) of a single unique BCR, which has the potential to recognise and bind antigen epitopes. Of the 10^{11} circulating B cells within an individual, the majority have unique antigen specificity. Antibodies are the secreted form of BCRs and are produced by plasma-cells forming part of humoral immunity (Reth et al., 2015). Antibodies can permeate extracellular spaces and recognise and bind the same antigen as their membrane bound counterpart, initiating a variety of downstream functions (Hoffman et al., 2016).

1.2.4.2 B Cell Receptor and Antibody Structure

BCRs and antibodies have two key functions, firstly to bind and recognise antigen and secondly to elicit a downstream response. This requires the coupled action of both a highly variable region capable of recognising the vast array of antigen encountered throughout life and a restricted constant region capable of binding the limited number of downstream receptors with high affinity. The evolutionary solution is the antibody, a glycosylated protein molecule comprised of four protein chains, two heavy and two light chains, linked together by disulphide bonds to form a Y-shaped molecule [**Figure 4**]. The N terminal region (Fab) is the antigen-binding site and contains linked heavy and light chains produced by V(D)J recombination to generate receptor diversity. Both 'arms' of the antibody possess identical antigen-binding regions and antigen specificity. This increases the avidity of the antigen-antibody interaction. The C terminal region (Fc) of the two heavy chains is known as the constant region. Constant regions are classified into five groups according to immunoglobulin isotype (IgM, IgD, IgG, IgA and IgE) and elicit the antibody isotype-specific effector function (Hoffman et al., 2016). Heavy chains are denoted by the corresponding lower-case Greek letter: μ, δ, γ, α, and ε respectively. There are two types of

light chains in humans, the kappa chain (ϰ) and the lambda chain (λ) (Collins and Watson, 2018). The genes encoding the light chains are found at two separate loci, unlike heavy chain constant genes.

Early in development B cells express both IgD and IgM, however the process of class-switch recombination (CSR) in response to antigen exposure, enables the expression of different immunoglobulin isotypes via DNA deletion (Stavnezer et al., 2008). Mature B cells express a single class of light chain (kappa or lambda), however instances of dual-expression in cancerous populations such as myeloma and B cell neoplasms are reported in the literature (Jiang et al., 2018, Xu, 2006b).

BCRs differ structurally from antibodies in two ways. They contain a hydrophobic domain on the heavy chain to anchor the receptor in the B cell membrane and a cytoplasmic tail which enables intracellular signal transduction with assistance from signalling proteins which form the BCR complex (Dylke et al., 2007).

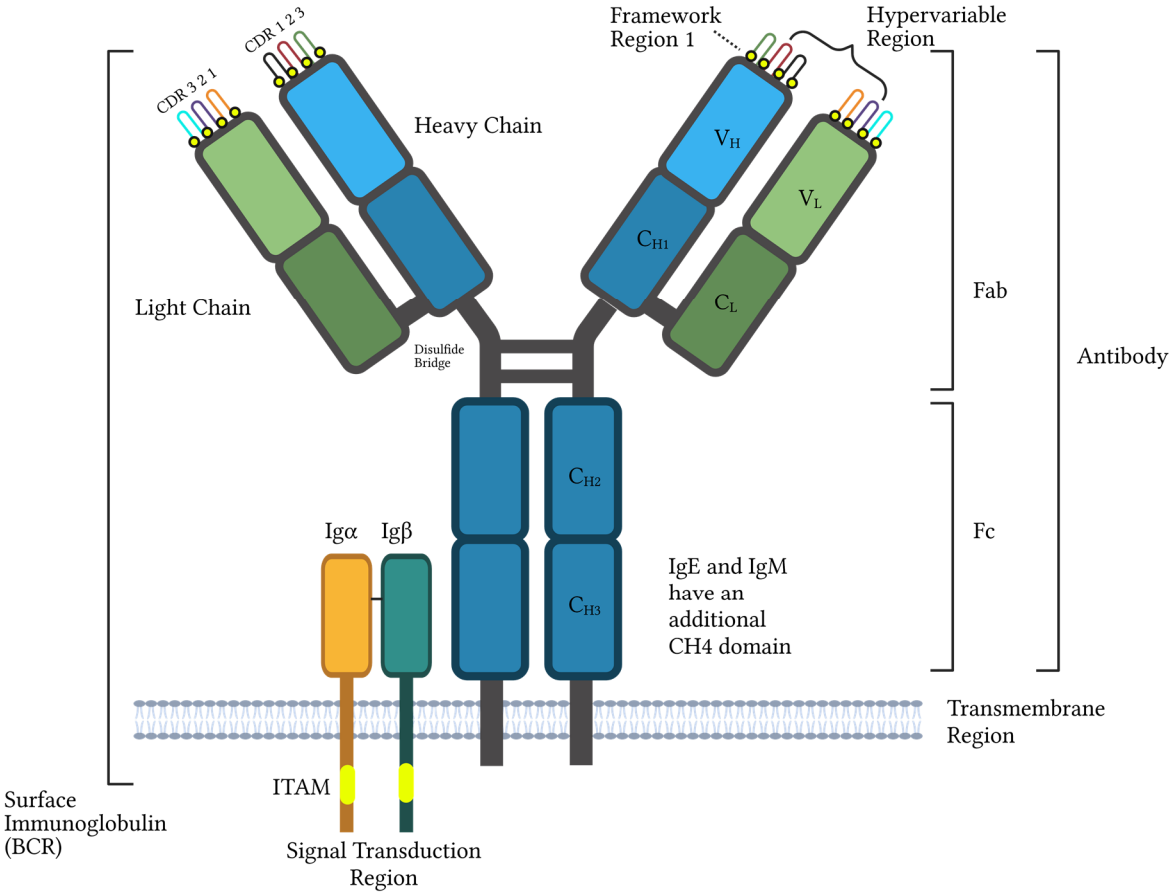


Figure 4. Representative structure of the human B cell receptor (BCR) and antibody.

BCRs consist of a membrane bound immunoglobulin that utilises four or five constant region domains, depending on immunoglobulin isotype. They also have a variable region responsible for antigen binding. V_H denotes the variable region generated by VDJ recombination of the Ig heavy chain locus while V_L represents the variable region generated by light chain VJ recombination. Each variable region is comprised of seven amino acid regions. This includes four relatively conserved framework regions (FRW) that provide structure. These regions are interspersed with three hypervariable regions, termed complementarity determining regions (CDR) 1, 2 and 3. The CDRs play a crucial role in antigen recognition and binding. The BCR is associated with an antigen non-specific signalling molecule called disulphide-linked Igαβ through non-covalent interactions. Igα and Igβ contain a single immunoreceptor tyrosine based-activation motif (ITAM) in their cytoplasmic domain. Upon antigen binding, the ITAMs become phosphorylated, initiating signalling pathways. Antibodies are the secreted form of the BCR. Figure adapted from Kovaltsuk et al. (2017).

1.2.4.3 Immunoglobulin Isotypes

There are five classes of immunoglobulin: IgM, IgD, IgE, IgA and IgG. IgG is divided into four subclasses (IgG1-4) and IgA is divided into two subclasses (IgA1-2). All classes of immunoglobulin can exist as either membrane bound BCRs or secreted antibodies, but they have distinct structures, distributions and downstream effector mechanisms. Despite

differences in amino acid sequences, individual immunoglobulin isotypes fold to form similar Y shaped monomers. However, IgA and IgM isotypes undergo dimerization and polymerisation resulting in complete antibody proteins with more complex structures [**Table 2**]. The downstream effect of immunoglobulins is predominantly mediated by Fc gamma receptors (FcγRs). There are six human FcγRs encoded by genes on chromosome one: CD64 (FcγRI), CD32a (FcγRIIA), CD32b (FcγRIIB), CD32c (FcγRIIC), CD16a (FcγRIIIA) and CD16b (FcγRIIIB) (Shankar-Hari et al., 2012).

Isotype	Subclass	Serum (%)	Serum Concentration	Structure	Complement Fixation	Opsonising	Cross Placenta	Other Functions	FcR
IgG	Total	75%	10 to 16 mg/mL	Monomer	+	+++	+	Secondary response. Neutralise toxins and Viruses.	FcγR
	IgG1	67% of IgG	6.98 mg/ml	Monomer	Yes	Yes	+		I,II,III
	IgG2	22% of IgG	3.8 mg/ml	Monomer	Yes	Yes	+		II
	IgG3	7% of IgG	0.51 mg/ml	Monomer	Yes	Yes	+		I,II,III
	IgG4	4% of IgG	0.56 mg/ml	Monomer	No	No	+		I,II
IgM	-	10%	0.5 to 2 mg/mL	Pentamer, Hexamer	+++	+	−	Primary response	Fcα/μR
IgA	Total	15%	1 to 4 mg/mL	Monomer, Dimer	−	−	−	Mucosal response, direct neutralisation of toxins, viruses and bacteria	FcαR (CD89)
	IgA1	85% of IgA	0.6-3 mg/ml	Monomer, Dimer	−	−	−		
	IgA2	15% of IgA	0.06-0.6 mg/ml	Monomer, Dimer	−	−	−		
IgD	-	<0.5%	0 to 0.4 mg/mL	Monomer	−	−	−	Homeostasis	FcδR
IgE	-	<0.01%	10 to 400 ng/mL	Monomer	−	−	−	Allergy	FcεR I,II

Table 2. Properties of immunoglobulin isotypes.

The major effector functions associated with each immunoglobulin isotype are indicated as '+++', while lesser functions are denoted by '+'. The absence of a corresponding function is indicated by the symbol '–'. Table adapted from Schroeder and Cavacini (2010), Vidarsson et al. (2014) and Furst (2009).

1.2.4.3.1 IgM

Monomeric IgM is the first immunoglobulin expressed during B cell development and exists on all developing B cells as the BCR. Mature antibody-secreting B cells secrete pentameric (occasionally hexameric) IgM, where single IgM units link together by disulphide bonds and polypeptide joining (J) chains to form a larger structure. Pentameric IgM is the first antibody to be secreted during the immune response and two classes have been identified: natural IgM and immune/induced IgM. Natural IgM antibodies are encoded by un-mutated germline variable gene segments and are constitutively secreted by plasma cells in the bone marrow and spleen, without stimulation by a foreign antigen. Natural IgM is associated with protection from systemic infections and constitutes the majority of secreted IgM found in normal serum, because of this it is considered part of the innate immune system (Blandino and Baumgarth, 2019). Typically, natural antibodies are polyreactive but have low affinity, although antigen binding is amplified by high avidity. Induced IgM is produced as a result of infection and these antibodies possess V region mutations indicative of post-germinal centre B cells which have undergone somatic hypermutation (SHM) and selection for increased receptor affinity. Post-germinal centre IgM B cells comprise 10-20% of all peripheral B cells (Keyt et al., 2020). Induced IgM enhances the humoral immune response and a sizeable population of non-class-switched memory B cells exists (Blandino and Baumgarth, 2019).

IgM predominantly functions by neutralisation, opsonisation and complement fixation, contributing to immune homeostasis by priming the adaptive immune response and supporting maximal IgG response. IgM can bind both a large number of altered self-antigens including debris from apoptotic cells and antigen produced during infection (Manson et al., 2005). For example, *Streptococcus pneumoniae* and oxidised LDL can be recognised by the

same natural IgM antibody (Binder et al., 2003). Where antigens contain repeating structures, pentameric IgM renders this process particularly efficient, hence IgM is especially effective against bacteria and viruses (Schroeder and Cavacini, 2010, Keyt et al., 2020). The formation of pentameric IgM enables two unique functions; it allows binding to lower-affinity antigen through strong avidity and it can be recognised and transported across cells by transcytosis delivering IgM to the mucosal epithelium. However, because of its size, pentameric IgM does not extravasate easily into tissues. When it does, for example due to endothelial damage, the presence of IgM in tissues causes excessive complement activation leading to tissue and organ damage (Matsumoto, 2022, Blandino and Baumgarth, 2019).

1.2.4.3.2 IgD

IgD circulates at low levels in the serum, has a short half-life due to its sensitivity to proteolysis and is the second least abundant isotype. The function of IgD is poorly understood and it is not known to partake in any of the major antibody effector functions. IgD processes are perhaps, mediated by an Fc receptor-independent interaction with basophils, mast cells and monocytes. Interestingly, IgD can bind certain bacterial proteins (such as from *Moraxella catarrhalis)* independently of the antibody variable region (Riesbeck and Nordstrom, 2006).

Recent evidence suggests IgD is strategically positioned for mucosal immunity as it is more readily detected in tonsillar mucosa and increases after infection with commensals and pathogens commonly found in the respiratory and middle ear mucosae including *Moraxella catarrhalis* and *Haemophilus influenzae*. A non-canonical class switching from IgM to IgD secreting B cells has been observed in human upper respiratory mucosa. In membrane-bound form, IgD is dually expressed with IgM in naïve B cells and may regulate B cell fate at

specific developmental stages (Schroeder and Cavacini, 2010). IgD increase relative to IgM is a reported feature of anergic B cells (Gutzeit et al., 2018).

1.2.4.3.3 IgG

IgG is the most abundant circulating antibody isotype and it has the longest half-life of all immunoglobulin isotypes. The four subclasses are highly conserved (>90% identical amino acid sequence) and numbered according to the circulating concentration in a western European population: IgG1>IgG2>IgG3>IgG4 (Vidarsson et al., 2014). Subclasses differ in antibody flexibility and affinity for downstream receptors due to differences in the hinge and upper CH2 domain. The main function of IgG is the neutralisation of toxins and viruses but with variable efficacy *e.g.* IgG3 neutralises HIV virus more effectively than IgG1 (Schroeder and Cavacini, 2010). All subclasses except IgG4 can bind and activate C1q, the first component of the complement pathway leading to deposition of C3b and opsonisation of pathogens (Vidarsson et al., 2014, Bournazos and Ravetch, 2017). IgG1 and IgG3 antibodies are mainly produced in response to protein antigens and are strong mediators of Fc-receptors (FcyR) and complement-mediated effector functions while response to bacterial capsular polysaccharides is restricted to IgG2. IgG4 is unique in that it has limited ability to activate effector cells, but it can form molecules with two different antigen binding sites via Fab-arm exchange resulting in bispecificity (Schuurman et al., 2001). The IgG repertoire can be skewed by infection and elevations in isotype subclass usage are associated with a variety of different diseases (Vidarsson et al., 2014, Engelhart et al., 2017). Selective subclass deficiencies are not usually detrimental but can increase susceptibility to specific classes of pathogen.

1.2.4.3.4 IgA

IgA exists as a monomer in the serum at higher levels than IgM, but considerably lower than IgG. In mucosal surfaces and secretions (saliva and breast milk) 'secretory' IgA exists as a dimer and is the most prevalent isotype. The abundance of IgA in secretions requires considerable energy to make and maintain. Daily production of IgA is greater than all other antibody classes combined (Woof and Kerr, 2006). IgA has a short tail and so it associates with one joining-chain (J-chain) to form dimeric IgA and is bound by the polymeric Ig-receptor (pIgR) ectodomain called secretary component (SC) which asymmetrically contacts both IgAs and the J-chain helping to stabilise and protect the antibody from degradation (Kumar Bharathkar et al., 2020). The two subclasses, IgA1 and IgA2, differ most noticeably in their hinge region, altering flexibility and antigen affinity. The truncated hinge in IgA2 leads to a reduced susceptibility to bacterial proteases compared to IgA1, perhaps explaining why IgA2 is more prevalent in mucosal secretions while in the serum 90% of IgA is IgA1 (Schroeder and Cavacini, 2010).

IgA is critical for passive protection of mucosal surfaces where levels of complement and leukocytes are low, acting through direct neutralisation and prevention of antigen-mucosal-surface binding (Woof and Kerr, 2006). The receptor pIgR, delivers IgA into mucosal secretions, but can also transport pIgA-containing immune complexes from the lamina propria across the epithelium and out into the lumen. This enables neutralisation of antigens originating in the diet, environment, gut microbiota and mucosal infections (Woof and Kerr, 2006). More recently it has been demonstrated that IgA can actively mediate immune responses by modulating the expression of cytokines such as TNF and IL-1β by human myeloid cells, in conjunction with the activity of pattern recognition receptors (PRR),

meaning it can induce inflammatory and immune-suppressive responses (Hansen et al., 2019).

1.2.4.3.5 IgE

IgE is a very potent immunoglobulin. It exists in the lowest concentration in the serum and has the shortest half-life. IgE binds strongly to the FcεRI receptor which is expressed on mast cells, basophils, Langerhans cells and eosinophils. IgE cross-linking signals cell degranulation which releases inflammatory mediators and causes an immediate hypersensitivity response, hence its association with hypersensitivity, allergic reactions and response to animal venoms and parasitic worm infections (Sutton et al., 2019, Schroeder and Cavacini, 2010). Most IgE exists cell-bound due to high receptor affinity, therefore only a small amount of antigen binding is required for a rapid reaction.

1.2.4.4 B cell Development

1.2.4.4.1 Early Development

B cells develop in the bone marrow from hematopoietic stem cells which undergo a series of differentiation and maturation stages before migrating to secondary lymphoid organs where they complete maturation forming either follicular or marginal zone B cells. At each stage, close contacts between B cell precursors and the bone marrow stroma are required to provide the necessary survival factors, including growth factors and cytokines, which ensure progression along the maturation pathway by positively and negatively regulating gene expression (Eibel et al., 2014) [**Figure 5**].

Early progenitor B cells (pro-B) are the first stage of the B cell lineage, and at this point cells undergo VDJ recombination of the immunoglobulin heavy chain loci to generate a unique IgH sequence**.** IGHD and IGHJ genes are rearranged first, typically on both alleles

simultaneously. Then, one IGH-DJ segment is imprecisely joined to an IGH-V gene segment on a single allele. At this point the B cell is known as a precursor (pre)-B cell. Pre-B cells express the pre-BCR complex formed from a recombined IgH chain coupled to a surrogate light chain and signal transducing subunits Ig-α and Ig-β. Pre-BCR signalling promotes survival and proliferation signals. Generation of a functional heavy chain prevents V-DJ rearrangement on the other allele, which then relocates to pericentromeric heterochromatin in a process termed allelic exclusion. In the event of an unproductive receptor resulting in no pre-BCR signalling, VDJ rearrangement on the second allele occurs (Vettermann and Schlissel, 2010).

Pre-B cells then undergo VJ recombination on the light chains. Typically IgK locus rearrangement occurs first and if neither allele is successful, IgL rearrangement is attempted. As there are two potential light chains (IgK and IgL) there are four opportunities to generate a functional receptor following the principles of allelic exclusion. Only B cells with fully functional IgH and IgL chains (K/L) survive to become immature B cells which express the integrated IgM BCR on the cell surface (Pieper et al., 2013).

Immature B cells leave the bone marrow and migrate to the spleen where they finalise development and differentiate into naïve follicular or marginal zone B cells. Mature antigen naïve B cells express multiple copies of the unique functional BCR and are able to produce both IgM and IgD by alternative splicing of μ or δ constant regions to the IgHJ exon respectively (Xu et al., 2012). Each B cell can proliferate to form a lineage of B cells expressing a clonal BCR in response to antigen encounter.

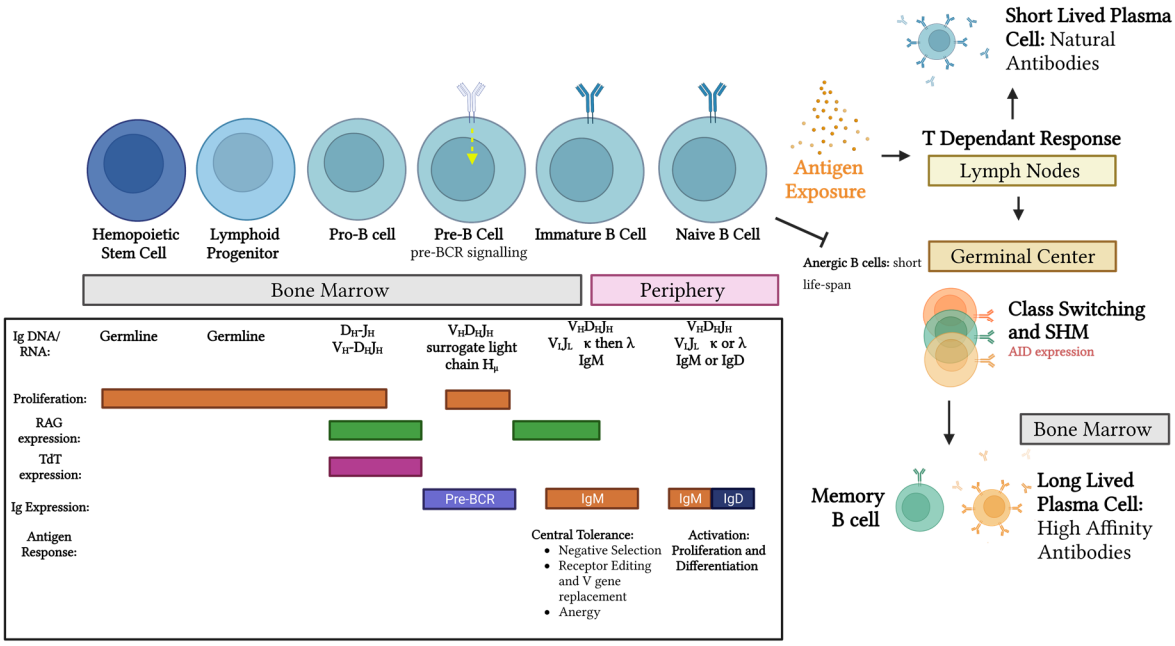


Figure 5. B cell maturation process.

Each B cell maturation stage is annotated with its corresponding anatomical site, surface immunoglobulin expression and BCR rearrangement status. The expression of key VDJ recombination proteins, namely the Rag complex and TdT, are also indicated. The expression of VDJ machinery is tightly controlled and occurs at specific stages of B cell differentiation. VDJ recombination will be discussed in section **1.3.1.1**. Figure adapted from Edwards and Cambridge (2006).

1.2.4.4.2 Central Tolerance

BCRs are produced by the imprecise joining of V, D and J gene segments (discussed in section **1.3.1)**, which may consequently generate autoreactive B cells. To prevent the deleterious effects of autoreactive clones, three methods of B cell tolerance are employed within the bone marrow: negative selection, receptor editing (and IGHV gene replacement) and B cell anergy (Barnas et al., 2021).

1.2.4.4.2.1 Negative Selection

Immature B cells with high affinity for self-antigens expressed in the bone marrow (autoreactive), receive a strong signal via the BCR signalling pathway. This signal arrests differentiation and the cell undergoes apoptosis within 2-3 days in a process called clonal deletion (Barnas et al., 2021). Clonal deletion is not the primary mechanism for preventing

autoreactive B cells, rather it occurs when another system (*e.g.* receptor editing) fails (Pelanda and Torres, 2012). Similarly, immature B cells expressing a non-functional BCR receive no pre-BCR signalling and apoptosis occurs.

1.2.4.4.2.2 Receptor Editing and V Gene Replacement

Receptor editing is the process whereby B cells reactivate the Ig rearrangement program and generate a new chain with different antigen specificity. Receptor editing separates cellular selection and BCR selection; it is possible to remove an autoreactive BCR without destroying the B cell. Editing normally occurs on the light chain and most often, all or part of the V and J gene is replaced with an upstream V gene and downstream J gene (Darlow and Stott, 2005). Receptor editing can rescue B cells committed to cell death due to autoreactivity or expression of a non-functional BCR, provided the new BCR has reduced or no autoreactivity. As receptor editing most often involves the light chain, it is particularly beneficial as in many cases considerable energy will have already been expended generating a functional IgH chain.

As the IgH chain is comprised of V, D and J gene segments and all but one D_H is deleted during VDJ recombination, in principle, receptor editing is not possible (Pelanda and Torres, 2012). However, most or all of the V_H gene in the existing VDJ exon can be replaced with an upstream V gene using a cryptic recombination signal; this is called V gene replacement. However, the probability of generating a functional receptor is low (Chen et al., 1995). V gene replacement can cause elongation of the CDR3 region and promote autoreactivity. It is identified by a characteristic 'footprint' left in the VDJ sequence (Luning Prak et al., 2011). V gene replacement predominantly occurs in pre-B cells that have not yet undergone light chain rearrangement (Wang et al., 2020b, Sun et al., 2015).

Emerging evidence has shown that receptor editing can also occur in mature B cells with evidence of SHM. Hence, BCRs underwent revision in the germinal centre, not as a mechanism to prevent auto reactivity early in development but rather to promote repertoire diversification in response to antigen encounter. In this context, the process is termed receptor revision and has been predominantly documented in large expanded clones, with high SHM, suggesting that it is a rare event which occurs in the presence of chronic antigen stimulation or auto-immunity (Nemazee and Weigert, 2000).

1.2.4.4.2.3 Anergy

B cells with intermediate binding of the BCR to self-antigens in the bone marrow can be functionally inactivated. These B cells are released to the periphery but are non-responsive (Pieper et al., 2013). Anergic B cells may be reactivated by non-BCR signalling pathways and contribute to autoimmune disease but they have a reduced lifespan in the periphery compared to non-autoreactive B cells, so are effectively deleted over time (Nemazee, 2017).

1.2.4.5 Response to Antigen

1.2.4.5.1 BCR Activation

Naïve B cell activation requires an intracellular signalling cascade initiated via BCR-antigen binding. Activation leads to the internalisation of antigen for processing and presentation to T cells. Naïve B cells express IgM and IgD which both lack a significant intracellular domain. Signal transduction relies on the Igα and Igβ complex which is pre-associated with IgM/D. BCR signalling regulates the expression of a number of downstream genes, including co-stimulatory molecules, adhesion molecules, migration receptors, pro-survival molecules and cell cycle related genes (Cyster and Allen, 2019). The mechanism underpinning BCR activation is debated and several hypotheses have been proposed [**Table 3**].

Model	Mechanism
Conformation-induced oligomerisation	Antigen binding causes cross linking of IgM monomers on the cell surface by unmasking the clustering interface of the Cμ4 domain using a pulling or twisting force. Rapid clustering of B cell receptors forms a membrane bound signalosome and opening of cytoplasmic domains initiates BCR signalling (Treanor, 2012, Cyster and Allen, 2019, Tolar et al., 2005, Liu et al., 2016a).
Dissociation activation	BCRs pre-exist on the cell membrane as auto-inhibitory oligomers/clusters. Antigen binding dissociates these clusters, exposing phosphorylation residues within Igα/Igβ co-receptors which enables signalling (Yang and Reth, 2010, Treanor, 2012).
Collision coupling	BCRs on the cell membrane are spatially separated from activation receptors or kinases. The motility of the BCR relative to its co-receptors is altered by antigen binding, enabling functional interaction and cell signalling (Treanor et al., 2010, Treanor, 2012).

Table 3. Hypothesised mechanisms of BCR activation in response to antigen binding.

B cell antigens are typically divided into two groups according to whether T cell stimulation is required for B cell activation; T dependent (TD) and T independent (TID) antigens. These groups differ in size and structure and the majority of antigens lead to TD responses. Although TD antigen cannot induce B cell proliferation without T cell help, they produce class-switched B cell responses with high antigen-affinity.

1.2.4.5.2 T Dependent Responses

In TD responses, antigen-engaged B cells internalise antigen, process it and return it to the cell surface as peptides bound to MHC class II molecules. B cells upregulate the chemokine receptor CCR7 which localises them to the T-B border of peripheral lymph nodes. Here they rely on long-lasting physical encounters with rare cognate T follicular helper (T_{fh}) cells specific to the presented peptide-MHC complex for activation. The matched T_{fh} cell delivers B cell survival signals including: CD40L, SLAM, SAP, ICOS, LFA-1, BAFF and the cytokines IL-4 and IL-21. This can trigger early class-switch recombination events (Cyster and Allen, 2019, Biram and Shulman, 2020). BCRs and matched TCRs do not recognise identical epitopes as processed peptides differ structurally to the surface epitopes recognised

by the B cells. T cells are pre-activated by co-stimulatory signals from encounters with professional antigen presenting cells such as dendritic cells (Biram and Shulman, 2020).

A small proportion of activated B cells differentiate into early, low-affinity, short-lived plasma cells and memory cells in the extra-follicular regions of secondary lymphoid organs. The remaining B cells migrate to the B cell follicular region of secondary lymphoid tissues, forming specialised micro-anatomical structures called germinal centres (GC). Here they undergo rapid rounds of proliferation and somatic hypermutation for BCR diversification, discussed further in **1.2.4.5.4**. GC B cells undergo positive selection for clones with high antigen affinity, expanding the antigen-reactive B cell pool and producing BCRs with higher/optimal binding affinity. GC B cells differentiate into either long-lived plasma cells which migrate to the bone marrow and secrete antibodies or memory B cells, primed for a secondary response (Martinez et al., 2021a).

1.2.4.5.3 T Independent Responses

In TID responses, antigen engages both BCRs and non-BCR surface receptors including toll-like receptors and complement receptors. This amplifies or modifies BCR signalling, bypassing the need for T_{fh} cell help. It may also reduce the threshold of antigen needed for B cell activation. TID antigens are diverse and divided into two classes: polyclonal B cell stimulants (soluble antigen, T_{indep} Type I) and large non-protein molecules with highly repetitive epitopes (T_{indep} Type II) which bind multiple BCRs at once. Many bacterial cell walls and viral capsids are Type II and trigger TID responses (Liao et al., 2017a, Vos et al., 2000).

1.2.4.5.4 B Cell Somatic Hypermutation and Germinal Centre Reaction

In T-dependent responses, upon antigen binding to the BCR, activated B cells migrate to the germinal centre (GC). The GC is divided into two distinct compartments, the dark zone and the light zone. In the dark zone, GC B cells (centroblasts) undergo rounds of rapid proliferation (clonal expansion) and somatic hyper mutation (SHM) of the BCR variable region to increase the size and diversity of the antigen-reactive B cell pool. Centroblasts then migrate to the light zone for selection where they compete to capture antigen presented on the surface of follicular dendritic cells (FDCs). Captured antigen is internalised, processed and presented to T_{fh} cells (Nouri and Kleinstein, 2020). B cells which increase their affinity for antigen via SHM are better able to capture the limited amount of available antigen and consequently receive greater levels of survival signals from T_{fh} cells which in turn, promotes further rounds of proliferation and selection. B cells which receive insufficient signal undergo apoptosis or are suppressed as part of immunological tolerance. Clonal affinity selection promotes the development and expansion of B cells with optimal binding properties. When the received signals (mediated by the BCR and co-stimulatory molecules) are above a certain threshold differentiation into long-lived high-affinity antibody-secreting plasma cells or memory B cells occurs. The threshold for differentiation into memory B cells is less stringent than for plasma cells (Stebegg et al., 2018, Gatto and Brink, 2010).

SHM introduces single nucleotide substitutions at a frequency of $\sim 10^{-3}$ per base pair per generation, a rate significantly higher than background mutation in non-immunoglobulin B cell genes (Pilzecker and Jacobs, 2019). Activation-induced cytidine deaminase (AID) is expressed specifically in activated B cells and initiates both SHM and class-switch recombination. AID preferentially binds single-stranded DNA from rearranged Ig genes and

deaminates cytosine, generating highly mutagenic deoxy-uracil (U) (Pilzecker and Jacobs, 2019). The mismatch U:G lesion is often excised by uracil-DNA glycosylase leaving an abasic site. Damage response pathways are then activated and error prone DNA synthesis over the U:G lesion or abasic site generates a spectrum of mutations including, C/G transitions, C/G transversions and A/T conversions around the initial lesion (Pilzecker and Jacobs, 2019). Mutation sites are non-random and preferentially occur at mutational hotspots, which often exist within a WRCY consensus sequence. BCR heavy and light chain variable regions contain three complementarity determining regions, CDR1, CDR2 and CDR3 which are hotspots for somatic hypermutation and are involved in antigen binding. The CDR3 region is the most diverse and the major determinant of antigen specificity as it is encoded by the junction of V, (D) and J gene segments (Nguyen et al., 2016). Germline IgV genes exhibit biased codon usage and CDR regions are intrinsically more mutable than the rest of the sequence. This, to a degree, localises SHM to regions of the antibody which are involved in antibody binding, while preserving the overall structure of the antibody by limiting mutations in framework regions (Di Noia and Neuberger, 2007).

1.2.4.5.5 Class-Switch Recombination

T_{fh} cells secrete cytokines which promote class-switch recombination (CSR) in activated IgM/D B cells, replacing the expressed immunoglobulin with a downstream isotype (IgE/A/G) altering antibody effector function (Stavnezer et al., 2008). Stimuli are classed as either primary CSR-inducing stimuli which upregulate the enzyme AID and proteins required for CSR machinery or secondary CSR-inducing stimuli which direct class switching by selecting the acceptor 'S' regions (upstream of each constant region exon), including interleukin-4, interferon-γ and transforming growth factor-β (Xu et al., 2012).

CSR occurs predominantly in GC B cells, following the binding of CD40 on B Cells with trimeric CD40 ligand (CD154) expressed by T_{fh} cells. T cell independent CSR can be induced by TLRs at very low levels early in infection (Xu et al., 2012). Switch (S) regions are comprised of tandem repeat units and flank the ordered array of IgH constant regions. Each constant (C) gene, except Cδ, is preceded by an S region. The donor S region (Sμ) is upstream of the Cμ gene, the acceptor S regions are upstream of the remaining C genes (Sɣ3, Sɣ1, Sα1, Sɣ2, Sɣ4, Sε, Sα2). During class switch recombination double strand breaks form at two participating switch (S) regions, determined by secondary CSR-inducing stimuli, on germline transcripts which cover the C_H exon cluster (Xu et al., 2012). Induction of AID loops the intervening DNA forming an extra-chromosomal DNA recombination product called the switch circle which is 'deleted' from the DNA. Re-join and repair of the broken ends through DNA repair pathways follows, producing an edited sequence that encodes for the class switched antibody isotype (Cyster and Allen, 2019).

The Ig constant region most proximal to the IGHV-D-J exon determines the antibody isotype, hence CSR has inherent directionality [**Figure 6**]. For example, as the IgE constant region is located downstream of IgG region, a B cell may undergo CSR from IgM/D to IgG and another CSR event can occur to IgE. However, an IgE B cell may not switch to IgG as the IgG constant regions were previously excised and cannot be brought proximal to the IgHV-D-J segment. IgA2 constant regions are the most distal and CSR to IgA2 is a terminal event (Cyster and Allen, 2019).

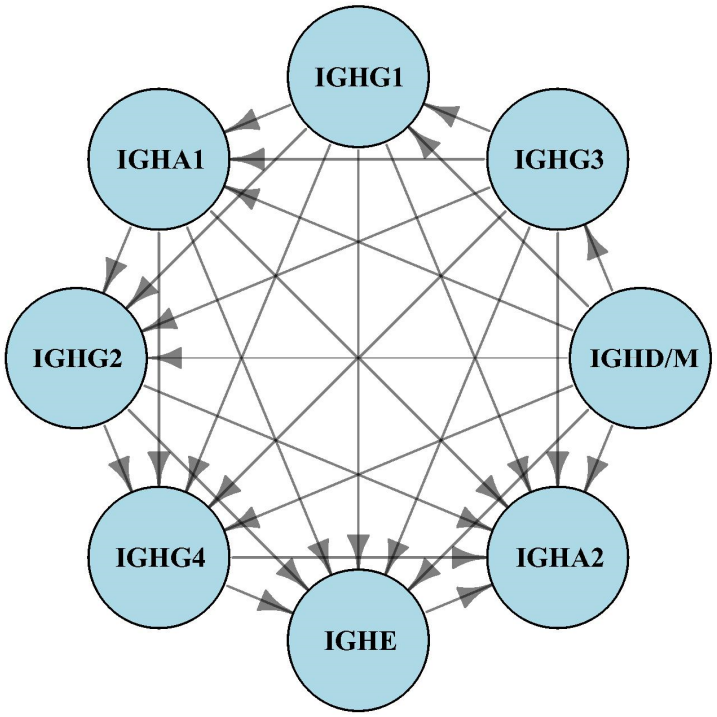


Figure 6. Class switch recombination network.

Illustrating sequential immunoglobulin isotype usage, through intra-chromosomal deletion recombination. The network is initiated with IGHD/M expression and terminated at IGHA2, the immunoglobulin isotype most proximal to the IGHV-D-J exon.

1.2.5 T Cell Biology

1.2.5.1 Introduction

T cells are a major component of the adaptive immune system and coordinate the immune response to pathogens, allergens and tumours. They are defined as those lymphocytes which express a T cell receptor (TCR) (Mariuzza et al., 2020). There are two T cell lineages which express different TCRs. Conventional αβ T cells express TCRs composed of α and β variable chains and recognise peptides presented on major histocompatibility complex (MHC) molecules. Most αβ T cells express either CD4 (30-60%) or CD8 (10-30%) which is important in MHC recognition. CD4 associates with MHC class II molecules which present endocytosed antigen peptides of 14-25 residues and CD8 associates with MHC class I which presents endogenous peptides of 8-10 residues (Rock et al., 2016).

T cells populate almost every organ and tissue in the body. The majority are found in lymphoid tissue, only 2-3% are found in the blood (Kumar et al., 2018). Unlike B cells, T cells do not undergo somatic hypermutation and selection upon antigen exposure therefore TCR antigen affinity is magnitudes lower than for BCRs (Alcover et al., 2018).

Approximately 5% of T cells express TCRs made from γ and δ chains. They are subdivided according to the type of TRVγ and TRVδ chain they express, which is associated not only with their function but also their tissue localisation [**Table 4**]. There are two major γδ T cell populations in humans, Vδ1 T cells are predominant in the thymus, human epithelia, skin, liver and spleen, while Vδ2 (in association with the Vγ9 chain) T cells constitute the majority of blood γδ T cells (Lawand et al., 2017). Around 70% of γδ T cells are $CD4^-CD8^-$ ~30% are $CD8^+CD4^-$ and only a small percentage (<1%) are $CD4^+CD8^-$. Generally γδ T cells are more numerous in mucosal compartments (Alcover et al., 2018, Jensen et al., 2008).

Gamma-delta T cells recognise a limited and distinct group of antigens, usually intact protein antigens and small phosphate or amine-containing compounds, independently of MHC presentation. Recognition occurs via the TCR, Natural Killer (NK) receptors and Pattern Recognition Receptors. They react immediately upon encountering antigen, a response resembling that of antibody recognition and can attack target cells directly via their cytotoxic activity or indirectly via activation of other immune cells (Lawand et al., 2017). Therefore, γδ T cells are often considered a bridge between innate and adaptive immunity (Allison and Garboczi, 2002, Santana and Esquivel-Guadarrama, 2006). $Vγ9^+Vδ2^+$ are traditionally regarded as the more 'innate-like' subset, and respond to endogenous and exogenous phosphoantigens using a semi-invariant oligoclonal repertoire. By comparison $Vδ1^+$ γδ T cells pair with a wider array of Vγ segments and behave more like adaptive, conventional αβ T cells, exhibiting antigen driven clonal expansion and

differentiation from naïve to effecter phenotypes (Clark and Thomas, 2020). Individuals with bacterial, parasitic and viral infections often have elevated γδ T cell count in the peripheral blood.

Vδ gene	**Paired Vγ gene**	**Distribution**
Vδ1	Vγ1, Vγ3, Vγ4, Vγ5, Vγ8, Vγ9	Peripheral blood, skin, gastrointestinal tract, spleen and liver
Vδ2	Vγ9	Peripheral blood
Vδ3	Vγ2, Vγ3	Peripheral blood, liver and gut epithelium

Table 4. Gamma-delta T cell subsets.

Gamma-delta T cell subtypes are classified according to TRGV and TRDV gene usage and their corresponding tissue localisation in humans is annotated. Table adapted from Sengupta (2019).

1.2.5.2 T Cell Receptor Structure

T cells express multiple copies of a single unique TCR [**Figure 7**]. Both γδ and αβ TCRs are the product of V(D)J rearrangement during thymic development and, like BCRs, are comprised of two chains linked by disulphide bonds. Both chains contain an immunoglobulin (Ig)-like extracellular variable V domain for antigen recognition, a constant C domain which associates with the CD3 complex for intracellular signalling, a membrane-proximal connecting peptide (CP), a transmembrane segment and a short cytoplasmic tail which does not contain an intracellular signalling domain. Signal transduction is achieved through the associated CD3 complex (Dong et al., 2019). Each subunit (CD3εγ, CD3εδ, and CD3ζζ), contains an extracellular immunoglobulin domain, a short CP, transmembrane segment and an intracellular immunoreceptor tyrosine-based activation motif (ITAM). TCR-pMHC binding induces phosphorylation of the ITAM motif for downstream signalling (Dong et al., 2019). The variable regions of TCRγδ and TCRαβ are structurally similar, however their

constant regions differ greatly in shape and charge leading to differences in the biochemical structure of each TCR complex (Morath and Schamel, 2020) [**Figure 7**].

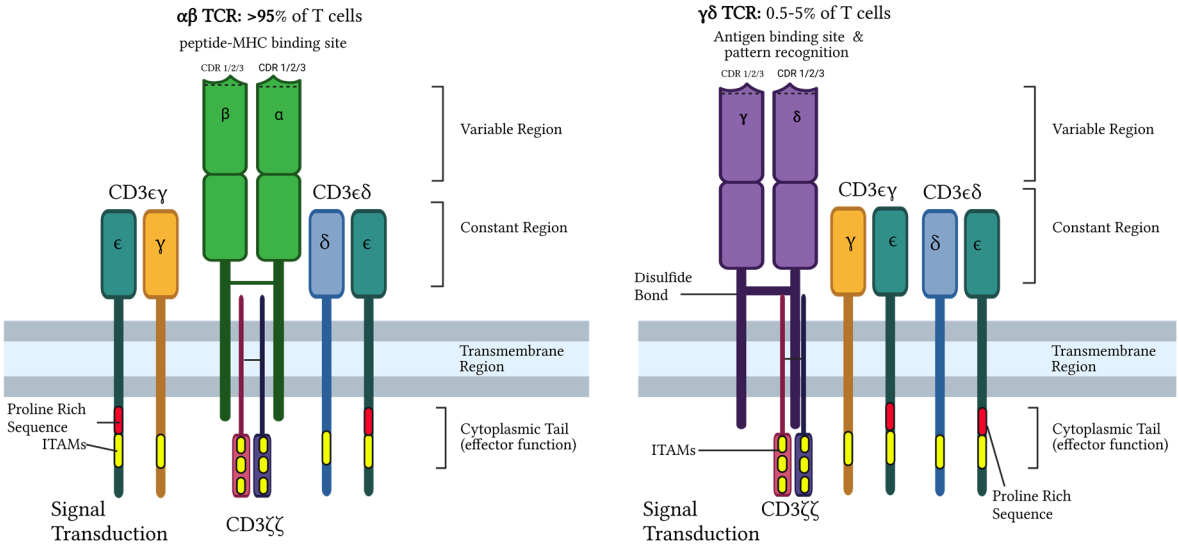


Figure 7. Representative structure of the human T cell receptor (TCR) complex.

TCRs are comprised of two chains (γ and δ or α and β) complexed with three dimers, CD3ϵγ, CD3ϵδ, and CD3ζζ. Each CD3 chain contains one or three intracellular immunoreceptor-tyrosine based activation motifs (ITAMs, yellow) that are crucial for signal transduction. Although there are similarities in the structure of TCRαβ and TCRγδ and in the shared CD3 subunits, the assembled geometry, glycosylation pattern, plasma membrane organisation and accessibility of signalling motifs in the CD3 intracellular tails differ between TCRαβ and TCRγδ complexes. Additionally, the distance between the immunoglobulin-like domains and the disulphide-bond in the CP is longer in TCRγδ compared to TCRαβ. Figure adapted from Morath and Schamel (2020).

1.2.5.3 Development

T cell development primarily occurs in the thymus from blood-borne progenitor cells which originate in the bone marrow. These progenitor cells can develop into T cells, natural NK cells, lymphoid and myeloid cells and lack expression of both CD8 and CD4; they are known as double negative (DN) thymocytes. DN thymocyte development can be divided into DN1, DN2, DN3 and DN4 stages based on differential CD44 and CD25 expression. After commitment to the T cell lineage, DN2 thymocytes undergo TCR-β, γ and δ loci V(D)J

rearrangement and are termed pre-T1 cells. DN3 is the critical stage that determines separation into the Tαβ and γδ lineages, up until this point DN cells are precursors to both lineages [**Figure 8**] (Spits, 2002, Qi et al., 2021).

1.2.5.3.1 αβ T Cell Development

T cells that generate a productive TCRβ chain undergo a process called 'β-selection', which promotes differentiation and commitment to the Tαβ lineage. B-selection is mediated by an effector sensor called the pre-TCR complex which is comprised of the TCRβ rearrangement product, an invariant pre-TCR α-chain and CD3 subunits. Productive TCRβ rearrangement enables ligand-independent signalling through the pre-TCR (Muro et al., 2019). In the absence of survival signals the cell undergoes apoptosis. Surviving cells proliferate and differentiate into $CD4^+CD8^+$ double positive (DP) cells and TCRα locus VJ rearrangement is initiated. Finally a complete TCRαβ/CD3 complex which is capable of recognising peptide-MHC complexes is expressed.

DP cells express a random rearrangement of VDJ gene segments and so the population includes those that are autoreactive, non-functional and those that bind and recognise antigen. T cells undergo both positive and negative selection, altering the T cell pool. DP cells are positively selected if there is low affinity interaction of the TCR and self-peptide-MHC complexes expressed on thymic epithelial cells. High affinity or no interaction triggers DP cell apoptosis by negative selection and null selection ('death by neglect') respectively. Surviving T cells differentiate into $CD4^+$ or $CD8^+$ single-positive (SP) mature T cells through two distinct steps, specification and commitment, driven by specific transcription factors that regulate the expression of lineage specific genes (Spits, 2002). SP cells then migrate to the medulla of the thymus, where they are screened for self-reactivity against pMHC displayed by medullary thymic epithelial cells, before dispersing to the periphery as

competent naïve T cells. If the interaction is strong, SP cells undergo negative selection or differentiate into regulatory T cells ensuring self-tolerance (Muro et al., 2019).

1.2.5.3.2 γδ T Cell Development

T cells which rearrange γ and δ chains prior to TCRβ, express the TCRγδ/CD3 complex on the plasma membrane. TCRγδ self-oligomerises and initiates unique intracellular signalling pathways, triggering 'γδ-selection' which confirms the generation of functional TCRγδ chains (Muro et al., 2019). 'γδ-selection' promotes differentiation into $CD5^{+}$ $CD24^{low}$ Tγδ committed cells. Unlike αβ T cells, γδ T cells do not undergo ligand-driven positive selection.

1.2.5.3.3 Lineage Commitment

The mechanism which determines T lineage fate is under debate. The 'signal strength model' proposes that a strong (potentially ligand-dependent) signal from a TCRγδ induced transgene drives Tγδ development via the ERK-Egr-Id3 pathway. If the signal is weakened, precursor cells instead give rise to the Tαβ lineage through ligand-independent pre-TCR signalling. However, studies in mice have shown that the number of ligand-specific γδ T cells is comparable between ligand-sufficient and deficient mice and encountering antigen in the thymus was neither required nor inhibitory in their development (Jensen et al., 2008). An alternative hypothesis, the stochastic-selective (pre-commitment) model, suggests that fate is determined prior to TCR rearrangement. TCRγδ or TCRαβ expression on T cell precursors simply confirms cell fate as mismatches result in apoptosis **[Figure 8]** (Qi et al., 2021).

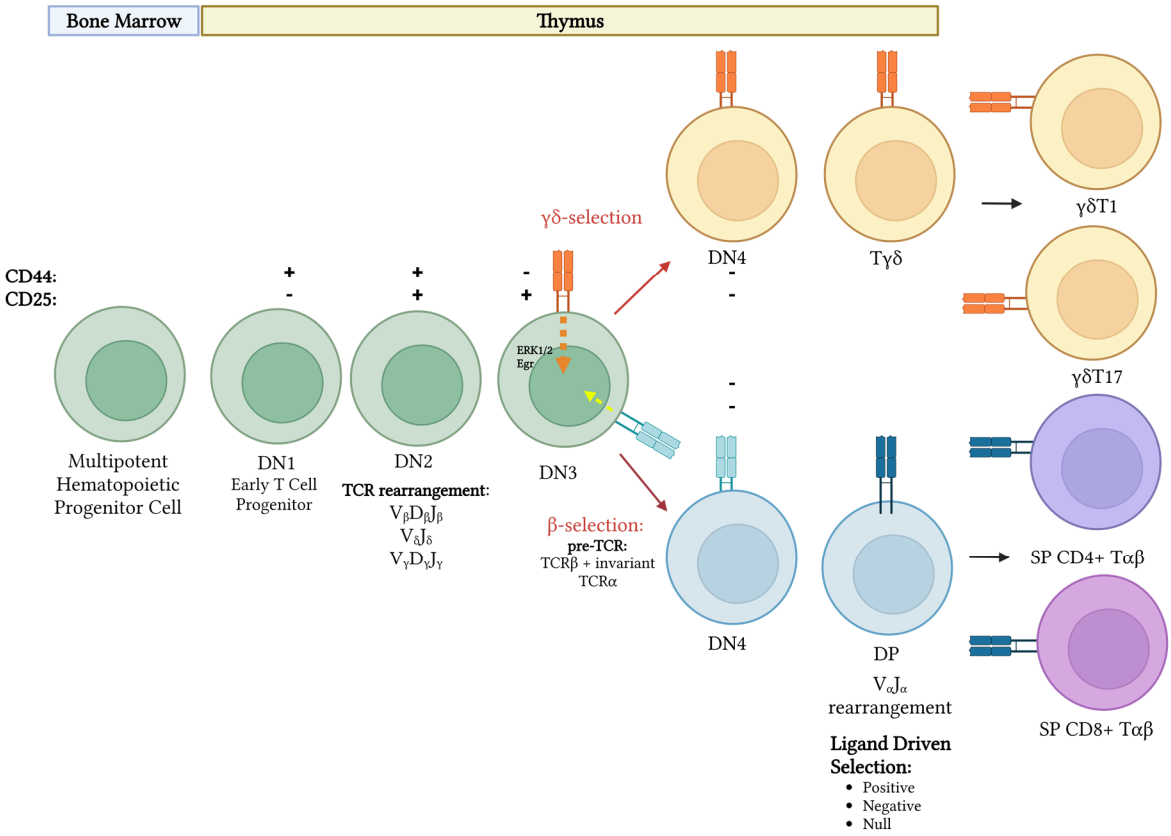


Figure 8. T cell development.

Schematic illustrating the development of γδ and αβ T cells in the thymus, starting from ancestral double negative (DN) thymocytes through to lineage commitment via γδ-selection and β-selection. Annotations include surface markers, VDJ rearrangement state, and the anatomical sites involved in each of the developmental stages. Additionally, the figure highlights the phenomenon of T cell polarisation, wherein γδ T cells can exhibit different phenotypes in response to various cytokines. For instance, a specific subset called γδT17 are pro-inflammatory regulatory T cells characterised by IL-17 production, while γδT1 cells produce both IFN-γ and TNF-α (Zhao et al., 2018). Adapted from Qi et al. (2021).

1.2.5.4 Response to Antigen:

1.2.5.4.1 αβ T Cell Activation

Alpha-beta T cell activation occurs when the TCR recognises processed antigen presented by the major histocompatibility complex (MHC) expressed on the surface of antigen presenting cells, infected host cells and tumour cells. TCR-pMHC engagement causes a conformational change in the CD3 complex leading to dissociation of the tyrosine-based activation motifs (ITAMs) from the inner leaflet of the T cell membrane (Dopfer et al.,

2014). ITAMs are tandem duplicates of a tyrosine containing sequence and each CD3γ/δ/ε subunit contains one ITAM, while the CD3ζ chain contains three. Protein tyrosine kinases (PTKs) phosphorylate the tyrosine residues within the ITAM, forming a signalling complex which interacts with effector molecules initiating downstream signalling (Hwang et al., 2020).

As with B cells, the mechanism by which TCR binding is translated into an activation signal in the CD3 complex is debated. The classic aggregation model proposes that clustering and cross-linking of TCRαβ/CD3 complexes with multimeric pMHC enables close contact and transphosphorylation between the CD3 tail and PTKs. However more recent studies have shown the presence of pre-existing TCRαβ aggregates on non-activated T cells (Smith-Garvin et al., 2009). The force-induced conformational change model suggests that mechanical force, generated by the sliding of T cell and antigen presenting cells membranes across each other, causes allosteric changes in TCRαβ conformation which propagates to the CD3 complex. This is amplified by TCRαβ-pMHC recognition (Mariuzza et al., 2020).

Activation of naïve T cells results in a proliferative burst, and individual cells are, in principle, capable of generating a large pool of functionally diverse progeny including short-lived effector T cells and long-lived memory T cells under the 'one cell, multiple fates hypothesis' (Gerlach et al., 2010, Buchholz et al., 2016). The strength and duration of the activation signal produced by TCRαβ stimulation, which is determined by TCRαβ affinity for the pMHC complex, is thought to specify naïve T cell fate. When naïve $CD4^+$ T cells receive strong TCRαβ stimulation they are more likely to differentiate into T_{h1} cells than T_{h2} cells. A weak signal has the reciprocal effect on $CD4^+$ T cells and in $CD8^+$ T cells is sufficient to generate $CD8^+$ memory T cells (Hwang et al., 2020). High-affinity TCRs, long-duration interaction and occupation of multiple ITAMs on a single CD3ζ drives the

differentiation into T_{fh} cells, which play an essential role in the formation of germinal centres and the production of high affinity B cells (Mesin et al., 2016).

1.2.5.4.1.1 T effector

The majority of naïve T cells in the periphery differentiate into effector T cells upon antigen binding to assist with pathogen clearance. Effector cells are short-lived and include $CD4^+$ helper, $CD8^+$ cytotoxic and $CD4^+$ induced (i) T_{reg} cells. $CD4^+$ T cells activate and regulate non-immune, B and cytotoxic T cells and help to suppress the immune response through the production of cytokines (Dong, 2021). $CD8^+$ T cells migrate throughout the body helping to clear infection by recognising p-MHC complexes on infected cells and inducing apoptosis. This is achieved through the interaction of surface molecules including Fas (CD95) and FasL (CD95L) and by the secretion of perforin and granzymes which cause membrane pore formation. $CD8^+$ T cells also release inflammatory cytokines which promote cell death (Schmidt and Varga, 2018).

1.2.5.4.1.2 T_{reg}

$CD4^+$ T_{reg} cells may also develop as a distinct lineage in the thymus and are termed natural T_{reg} (nT_{reg}) cells. Collectively, iT_{reg} and nT_{reg} cells comprise less than 5% of the T cell population in adults (Kumar et al., 2018). They maintain homeostasis and restrict activation of conventional T cells by limiting excessive immune response and suppressing most other immune cell populations, promoting tolerance to food and autoantigens. Binding of the TCR on T_{reg} cells to pMHC on dendritic cells (DC) results in either an antigen-specific tolerogenic dendritic cell, or prevents the DC from presenting a specific antigen (Shevyrev and Tereshchenko, 2019).

1.2.5.4.1.3 Memory T cells

Memory T cells are long-lived and persist post-infection after effector T cells have undergone apoptosis. They enable rapid response to secondary infection (Farber et al., 2014). Memory T cells potentially arise independently of effector cells from the common naïve progenitor, or they may derive from effector cells which escape cell death. $CD8^+$ memory T cells are poised to immediately proliferate, execute cytotoxic functions and secrete effector cytokines upon reinfection with an antigen (Martin and Badovinac, 2018). Memory $CD4^+$ T cells retain characteristics of effector $CD4^+$ T cells and can rapidly re-express effector molecules and release cytokines including (IFN)-γ and interleukin (IL-17) (Gray et al., 2018).

1.2.5.4.2 γδ T Cell Activation

Gamma-delta T cells are potent effector cells which recognise soluble or membrane-bound proteins via the TCR independently of MHC molecules. There is no apparent requirement for co-receptor interaction, suggesting that ligand binding is sufficient for activation (Deseke and Prinz, 2020, Dopfer et al., 2014). They recognise a variety of antigen including those produced by mycobacteria, plasmodium and cytomegalovirus infections and some self-antigens including stress induced and tumour antigens (Deseke and Prinz, 2020). Upon activation γδ T cells produce large amounts of T_{h1} cytokines *e.g.* IFN-γ and tumour necrosis factor (TNF)-α, triggering maturation of monocyte derived dendritic cells that act as antigen presenting cells which may enhance αβ T cell response (Pistoia et al., 2018). They also have the potential to be cytotoxic due to expression of surface receptors including natural killer group 2D and release soluble mediators (*e.g.* perforin and granzymes) similar to $CD8^+$ T cells (Lawand et al., 2017). There is some evidence that γδ T cells form long-lived memory T cells that reactivate upon reinfection and resemble memory αβ T cells in phenotype and recall-like response (Comeau et al., 2020).

1.3 The Adaptive Immune Repertoire

The adaptive immune repertoire describes the population of BCRs and TCRs expressed in an individual. BCR and TCR repertoires are complex and highly diverse, which enables high affinity immune responses to the wide variety of antigens encountered throughout life. BCR and TCR diversity is generated by V(D)J recombination, exonuclease trimming and the insertion of non-template nucleotides. B cells also undergo somatic hypermutation (Jackson et al., 2013). The location of BCR and TCR germline loci is detailed in **Table 5**.

Gene	Location	Number of Gene Families
IGH	14q32.33	IGHV:7, IGHD:7, IGHJ:6
IGK	2p11.2	IGKV:7, IGKJ:5
IGL	22q11.2	IGLV:11, IGLJ:7
TRA	14q11.2	TRAV:45, TRAVJ:61
TRB	7q34	TRBV:33, TRBD:2, TRBJ:2
TRG	7p14	TRGV:6, TRGJ:5
TRD	14q11.2	TRDV:3, TRDD:3, TRDJ:4

Table 5. BCR and TCR loci.

Table detailing the genomic locations and the number of V, D and J gene families for the human BCR and TCRs, as documented on IMGT (up-to-date as of January 2023).

1.3.1 Generating Diversity

1.3.1.1 V(D)J Recombination

The selection of one V, (D) and J gene segment and the generation of a functional V(D)J sequence in developing B and T cells determines the antigen specificity of each lymphocyte. This occurs via several recombination events in a process termed V(D)J recombination [**Figure 9**] (Roth, 2014, Schatz and Ji, 2011). V(D)J recombination follows 3 tightly-regulated phases. These are: 1) recognition of recombination signal sequences (RSS), 2) induction of double-stranded breaks and 3) repair of the broken DNA in a recombined configuration by components of the classical non-homologous end joining repair pathway.

The result is the deletion or inversion of a segment of DNA and the joining of V, (D) and J gene segments (Schatz and Ji, 2011). Errors can result in non-functional receptors.

BCR IgH and TCRβ/δ germline loci are comprised of multiple copies of variable (V), diversity (D) and joining (J) gene segments [**Table 6**], while BCR IgL and TCRα/γ loci are comprised of V and J gene segments only. This arrangement is maintained in most somatic cell types. As TCRα/γ and IgL loci only encode V and J genes, VJ recombination in these chains generates less diversity than VDJ recombination in their paired chain.

TCRγδ loci contain fewer total V, D and J gene segments than TCRαβ loci, therefore less diversity is generated during V(D)J recombination. To compensate, the TCRδ chain can increase diversity by utilising multiple copies of TRDD genes (Allison and Garboczi, 2002). TRDV gene segments are located between the TRAV and TRAJ genes, hence some TRAV segments are utilised by the TCRδ locus (Schatz and Ji, 2011). Sequence similarity between TRBV and TRDV genes and TRAV and TRGV genes indicate that they are evolutionarily related (Allison and Garboczi, 2002).

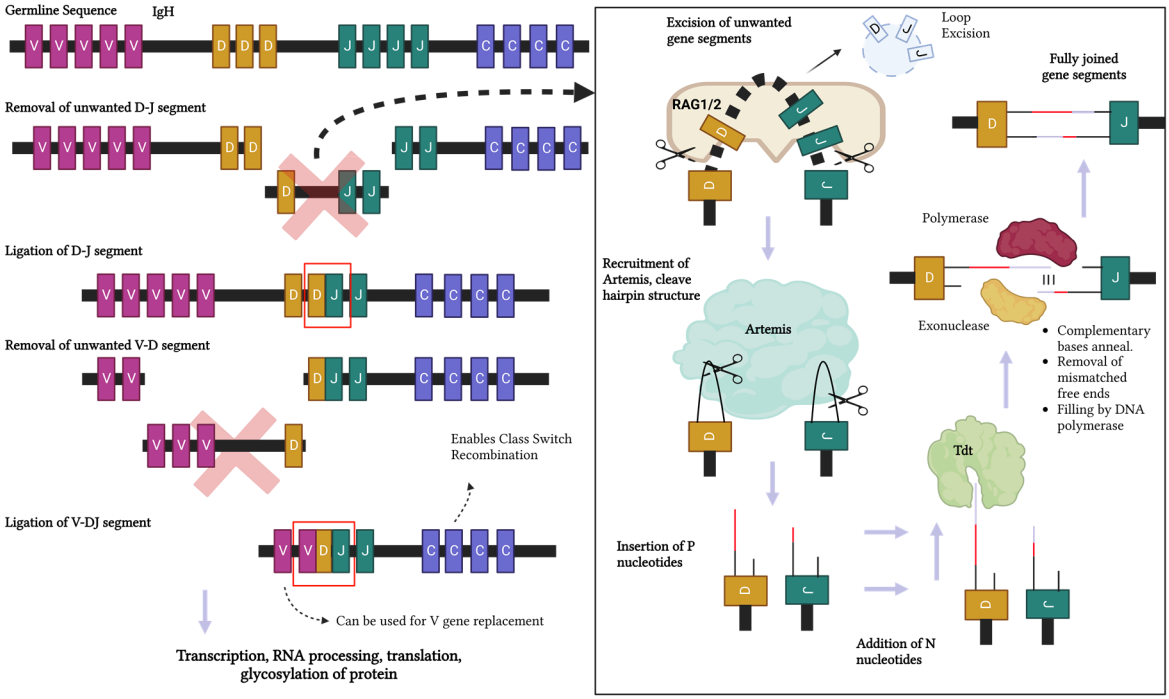


Figure 9. IgH chain VDJ recombination.

The schematic illustrates how VDJ gene rearrangement produces a functional B cell receptor (BCR). The use of different V, D, and J gene segments generates diversity in the BCR repertoire in a process termed combinatorial diversity. On the right side, the figure shows how junctional diversity, which arises from imprecise joining of gene segments and the incorporation of P (palindromic) and N (non-template) nucleotides functions to increase diversity. Adapted from Roth (2014) and Backhaus (2018).

Chain	Segment	Type	Functional	Non Functional	Total	Functional Combinations (Excluding Alleles)
IGH	**IGHV**	Gene	60	51	111	**8,280**
		Allele	309	129	438	
	IGHD	Gene	23	4	27	
		Allele	30	4	34	
	IGHJ	Gene	6	3	9	
		Allele	13	6	19	
IGK	**IGKV**	Gene	41	36	77	**205**
		Allele	72	48	120	
	IGKJ	Gene	5	0	5	
		Allele	9	0	9	
IGL	**IGLV**	Gene	33	46	79	**165**
		Allele	81	74	155	
	IGLJ	Gene	5	2	7	
		Allele	7	3	10	
TRA	**TRAV**	Gene	45	16	61	**2,295**
		Allele	109	26	135	
	TRAJ	Gene	51	10	61	
		Allele	60	11	71	
TRB	**TRBV**	Gene	49	19	68	**1,274**
		Allele	115	37	152	
	TRBD	Gene	2	0	2	
		Allele	3	0	3	
	TRBJ	Gene	13	1	14	
		Allele	14	2	16	
TRG	**TRGV**	Gene	6	9	15	**30**
		Allele	11	14	25	
	TRGJ	Gene	5	0	5	
		Allele	6	0	6	
TRD	**TRDV**	Gene	8	0	8	**96**
		Allele	24	1	25	
	TRDJ	Gene	4	0	4	
		Allele	4	0	4	
	TRDD	Gene	3	0	3	
		Allele	3	0	3	

Table 6. The number of functional and non-functional V, D and J gene segments for the TCR and BCR.

Data is sourced from the IMGT gene tables and is up-to-date as of July 2022 (source: https://www.imgt.org/). The expected number of functional VDJ combinations was calculated based on functional gene count.

1.3.1.1.1 Recognition of Recombination Signal Sequences

Each V, (D), and J gene segment is flanked by a recombination signal sequence (RSS) which consists of a conserved heptamer and nonamer element, separated by a 12 or 23 nucleotide spacer sequence. The differing spacer length is important as efficient recombination occurs between RSSs with different spacer lengths, known as the '12/23' rule (Schatz and Ji, 2011). The orientation of the RSS sequence determines whether recombination occurs due to deletion or inversion (Roth, 2014). In the IgH loci, the 3' end of V segments and the 5' end of J segments are flanked by 23bp RSSs, whereas the D segment is flanked on both sides by 12bp RSSs. This ensures that D segments interact with both J and V segments, but V and J segments cannot be joined. In the TCRβ/δ loci, the D gene segment is flanked by a 5'-12bp RSS and a 3'-23bp RSS implying that regulation of V-J joining extends beyond the 12/23 rule. In the light chain and TCRα/γ, the 3' end of the V segment and the 5' end of the J segment are reverse complementary, with a 12bp or 23bp spacer sequence allowing for VJ linkage (Chi et al., 2020). D-J recombination occurs first (where applicable) followed by V-(D)J recombination.

1.3.1.1.2 Induction of Double-strand Breaks

Recombination begins when two lymphocyte specific proteins known as recombination activating genes (RAG) 1 and 2 form a tetrameric complex called RAG recombinase. RAG1 contains the principle binding component and active site residues which catalyse DNA cleavage while RAG2 mediates and enhances chromatin binding (Christie et al., 2022). RAG recombinase recognises a single 12bp or 23bp RSS sequence and introduces a nick between the coding sequence and the RSS. Another RSS is then captured to form the paired complex and nicking follows at the second RSS. Transesterification by the free ends results in the formation of a hairpin complex. The endonuclease Artemis opens the hairpin, often

asymmetrically, resulting in a 3' nucleotide overhang which is the substrate for Terminal deoxynucleotidyl transferase (TdT) polymerase. Resolution of the overhang by DNA repair enzymes means that palindromic (P) nucleotides are often found in the final V(D)J sequence (Jackson et al., 2004). TdT polymerase can also add non-template (N) nucleotides to the overhang, causing 3'overhang elongation. This introduces further diversity into the V(D)J sequence (junctional diversity) and occurs more frequently in TCRα junctions (Jackson et al., 2013). Artemis then resects damaged DNA or non-ligatable end groups which can shorten the CDR3 region. Nucleotides can also be removed by exonuclease trimming, which is influenced by the nucleotide composition of the gene (Jackson et al., 2004).

1.3.1.1.3 Repair of Double-stranded Breaks

The two RSS (coding) ends of the hairpin (which may have diffused apart) are brought together in a synaptic complex for pairing and repaired quickly via the non-homologous end joining repair pathway (Christie et al., 2022). Any unpaired nucleotides are removed via an exonuclease. DNA polymerase then fills in the gaps by DNA synthesis and ligation to form the coding joint (Chi et al., 2020). The signal ends either side of the excised gene segments are blunt and held together by the RAG complexes. They are joined slowly compared to the RSS ends, forming a signal joint within a stable circular excision product.

1.3.2 Repertoire Analysis Methods

1.3.2.1 Low-Throughput Methods

Prior to the advent of Next Generation Sequencing (NGS), Sanger sequencing was used to sequence adaptive immune repertoires. Sanger sequencing is a low-throughput method and therefore fails to capture a complete representation of the diversity of the adaptive immune repertoire. The number of sequences captured is typically no more than a few hundred, so

it was commonly used to target specific clonal expansions identified by other technologies (Liu and Wu, 2018, Blachly et al., 2015, Six et al., 2013).

Several non-sequencing based approaches have been used to study immune repertoires. TCR and BCR spectratyping is a molecular based technique which quantifies CDR3 length distributions providing a measure of repertoire diversity and a means to identify clonal expansions. The CDR3 region is amplified from RNA by V-J gene specific reverse transcriptase PCR and CDR3 amplicons are size-separated using gel electrophoresis or capillary electrophoresis. Spectratyping is fast, low-cost and high-throughput, but nucleotide composition is lost limiting its utility (Ciupe et al., 2013, Liu and Wu, 2018). An up-to-date review of TCR spectratyping studies was published by Fozza et al. (2017).

Flow cytometric immunophenotyping (FCI) is a non-sequencing-based approach used, often in conjunction with spectratyping, to quantify V gene usage in αβ T cells. Cells are stained with a panel of fluorescently tagged V gene specific monoclonal antibodies and then sorted and quantified. FCI can identify V gene usage biases indicative of clonal expansion but fails to capture the full diversity of the TCR repertoire as sequence information is lost (Tembhare et al., 2011, Pilch et al., 2002).

PCR has been used to characterise the adaptive immune repertoire in multiple different ways. The simplest approach uses a panel of V gene specific primers to amplify the V-J or V-C region from DNA or RNA respectively. The relative concentration of amplicons are compared, for example by Southern blot analysis, to characterise V gene usage (Hase et al., 2000, Venet et al., 2013). PCR-based methods are susceptible to amplification bias and cannot capture diversity due to SHM and imprecise joining. Matsutani et al. (1997) developed a more complex approach using adaptor-ligation PCR to reduce bias in TCRβ V

gene amplification by excluding the need for multiple V gene primers. Amplicons are then quantified by hybridisation to a microplate coated with V gene specific oligonucleotides (Matsutani et al., 1997).

1.3.2.2 High-Throughput Methods

TCR V(D)J recombination can generate a theoretical maximum of 10^{15}-10^{20} different sequences, although the number of unique sequences expressed within an individual is estimated to be around 10^{13} (Rosati et al., 2017). BCR maximum sequence diversity is estimated to be $>10^{14}$ sequences differing by as little as one base pair due to SHM, however there are only 10^{10}–10^{11} B cells in human adults (Yaari and Kleinstein, 2015). Currently, NGS using the MiSeq system typically yields $>10^{7}$ reads per lane of sufficient length to cover the whole VDJ-constant region, so it is an ideal technique for repertoire analysis.

Both RNA and DNA can be used as starting material, although RNA is favoured as it contains the final TCR or BCR product without introns and non-functional or allelic excluded VDJ rearrangements. However, the presence of multiple TCR/BCR RNA transcripts per cell makes identifying individual clones challenging and can cause over-representation of certain sequences in the repertoire. For example, plasmablasts have a higher expression of IgH/IgL RNA compared to non-antibody secreting B cells and their BCR/antibody sequences are over-represented in repertoire data compared to what would be expected based on cellular proportions in blood (Rosati et al., 2017).

Before sequencing, most methods require targeted enrichment of the repertoire which is achieved in three main ways: 1) multiplex PCR, 2) 5'-rapid amplification of cDNA ends (5'RACE) and 3) RNA-capture. More recently, several computational methods have been

published which reconstruct VDJ sequences from total RNA sequencing data, bypassing the need for targeted amplification [**Figure 10**] (Song et al., 2021, Bolotin et al., 2017).

1.3.2.2.1 Multiplex PCR

Multiplex PCR followed by NGS is widely used as it is easy to perform, fast, highly specific and requires little starting material (*e.g.* RNA from single cells) compared to other enrichment methods (Peng et al., 2015). Commercial kits are available including: NEBNext Immune Sequencing Kit (New England Biolabs) and AmpliSeq for Illumina Immune Repertoire Plus (illumina). However, many studies preferentially use publicly available primer sets such as the BIOMED-2 primers or design and validate their own for flexibility in experimental design and reduced costs (van Dongen et al., 2003, Ly and Cotta, 2013). Unique molecular identifiers (UMIs) can be incorporated into amplicons during reverse transcription which enables computational correction of PCR amplification bias; one of the key limitations of pre-NGS studies. Some residual bias may persist if primer binding affinities are very different, hence primers must be carefully optimised before use (Peng et al., 2015). One major limitation of multiplex PCR is the pre-defined panel of V gene primers, which can prevent the detection of novel V alleles/variants (Rosati et al., 2017).

1.3.2.2.2 RNA-capture

RNA-capture was developed for human exome sequencing and uses RNA baits complementary to the sequence of interest to hybridise gDNA/cDNA which has been library-prepped. TCR and BCR loci can be targeted simultaneously (Mulder et al., 2018). Baits are bound to magnetic streptavidin beads which are used to retrieve captured DNA. Enriched DNA is submitted for universal amplification and sequencing (Linnemann et al., 2013). Fewer PCR cycles are required compared to multiplex PCR, reducing amplification bias. However, this relies on extensive bait panels which may be less efficient for capturing

sequences that have undergone SHM, and requires a high sequencing depth due to non-uniform DNA capture. RNA capture requires expensive hardware and large studies are currently unfeasible; a single person could realistically only perform 24 arrays a day (Mamanova et al., 2010).

1.3.2.2.3 5'RACE cDNA Synthesis and Nested PCR

5'RACE enables unbiased targeted enrichment of BCR/TCR sequences and detection of novel V genes. Constant region primers are used for first strand cDNA synthesis and a reverse transcriptase enzyme incorporates non-template nucleotides at the 3' end. A sequence-independent template switching primer anchors to the non-template sequence and the reverse transcriptase switches template and continues replicating to the end of the oligonucleotide for second strand cDNA synthesis. Next, PCR uses a constant region primer and a template switching primer for unbiased amplification of the VDJ region irrespective of V gene usage (Rosati et al., 2017).

5'RACE is inefficient and often generates incomplete sequences and RNA molecules with long untranslated regions will be lost. Fang, et al. report that only ~20% of TCRβ sequences generated by 5'RACE cover the full VDJ-C region (Fang et al., 2014, Lin et al., 2020). Bashford-Rogers et al. (2014) found that for a BCR repertoire, only ~55% of sequences generated by 5'RACE passed filtering compared to 76% using multiplex PCR with the same sequencing technology (MiSeq). Degraded RNA and short sequences from incomplete cDNA synthesis are easily polluted by sequencing adapters and a greater sequencing depth is required to account for this. Repeated sequencing of the same sample by 5'RACE is less reproducible than multiplex PCR when analysing shared CDR3 sequences (Liu et al., 2016b).

1.3.2.2.4 Bulk RNA Sequencing

BCR and TCR transcripts often occur at sufficient levels in bulk RNA sequencing data to characterise the adaptive immune repertoire. However, the repetitiveness and diversity of the loci means that the traditional mapping tools employed in RNA-seq analysis are ill-equipped to handle these sequences. In recent years, several computational tools have emerged which attempt to reconstruct complete VDJ sequences from bulk RNA-seq, negating the need for amplification of the BCR/TCR repertoire. These include: MiXCR, TRUST4 and ImReP (Song et al., 2021, Mandric et al., 2020, Bolotin et al., 2017, Bolotin et al., 2015). The performance of these tools varies across tissue and datasets and effectiveness depends on read length. Currently there is no consensus analysis pipeline.

Successful repertoire reconstruction paves the way for re-analysis of large public RNA sequencing databases like The Cancer Genome Atlas which contains >10,000 tumour samples (Mandric et al., 2020). As an additional benefit, paired repertoire and RNA-seq data enables comparisons between the repertoire and mRNA expression of every cellular gene such as those involved in T cell activation. Furthermore, the relative proportions of BCR/TCR genes are retained (Bai et al., 2018, Peng et al., 2022, Andreani et al., 2022, Smakaj et al., 2020). In T/B cell depleted tissues such as tumour biopsies, RNA-seq repertoire reconstruction only captures dominant high-frequency clonotypes, meaning tools such as these do not negate the need for targeted approaches which profile the adaptive immune repertoire at a greater depth (Peng et al., 2022). Instead, they can be a useful starting point where bulk RNA-seq data exists and there is pre-existing evidence of expanded repertoires. Targeted approaches could then be used where characterisation of less abundant clones is required (Bai et al., 2018).

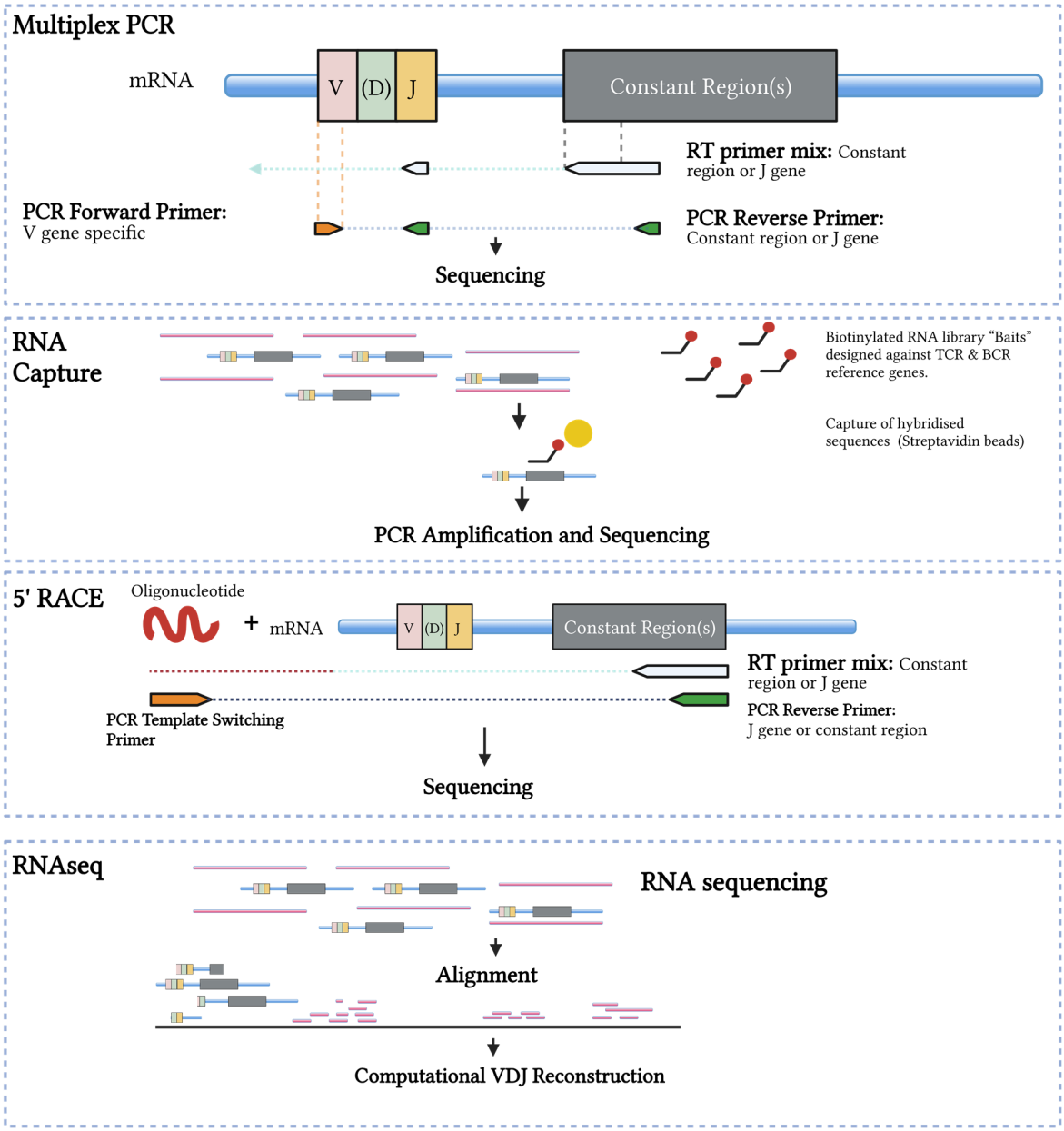


Figure 10. High-throughput BCR and TCR repertoire amplification methods.

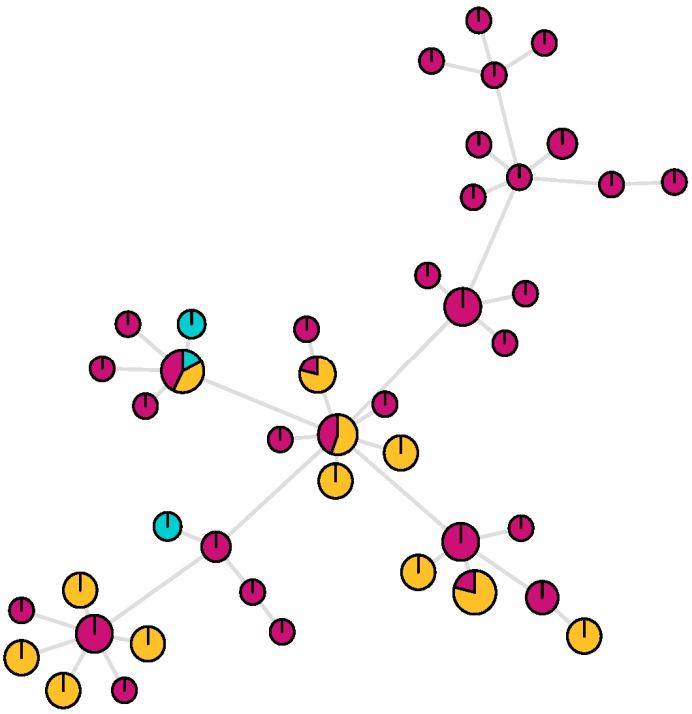
Schematic showing the experimental approaches for four high-throughput repertoire enrichment methodologies: multiplex PCR, RNA capture, 5' RACE and computational repertoire reconstruction using bulk RNA sequencing data.

1.4 Chapter 1 Aims

The overarching aim of this thesis is to characterise the adaptive immune repertoire in sepsis. Specific objectives are described below and will be further elaborated on in each chapter.

1) To amplify and sequence the complete adaptive immune repertoire (BCR IgH, TCR α, β, δ and γ chains) using an optimised in-house multiplex PCR reaction, for a cohort of patients with sepsis due to community-acquired pneumonia (CAP) and matched healthy controls. Additionally, to provide comprehensive guidelines on the pre-processing and quality control analysis for BCR/TCR repertoire NGS sequencing data.
2) To perform whole blood leukocyte bulk RNA sequencing on the same cohort of sepsis patients in order to assign SRS group.
3) To develop a novel computational approach for dimensional reduction of adaptive immune repertoire data, based on the principles of weighted gene correlation network analysis (WGCNA). This method will be compared to an alternative machine learning approach, sparse decomposition of arrays (SDA).
4) To characterise the BCR repertoire and TCR repertoire in sepsis focusing on features associated with mortality and clinical variables. Furthermore, to understand adaptive immune repertoire regulation by integrating repertoire modules with matched bulk RNA sequencing data (aim 2) using quantitative trait analysis.

Chapter 2. Materials and Methods



2 Materials and Methods

This chapter describes patient recruitment, BCR and TCR repertoire experimental approaches and the computational pre-processing pipeline. Bulk RNA sequencing and data processing will also be discussed.

2 Materials and Methods 80

2.1 Sample Selection 80

2.2 BCR/TCR Multiplex PCR Repertoire Amplification 84

2.3 Bulk RNA Sequencing 95

2.4 Computational Analysis 96

2.5 Statistical Analysis 115

2.6 B Cell Clonal Dynamics with MRDARCY 119

***A* Appendices: Materials and Methods 382**

2.1 Sample Selection

2.1.1 UK Genomic Advances in Sepsis Study

The UK Genomic Advances in Sepsis (GAinS) (ClinicalTrials.gov Identifier: NCT00131196) study was established in 2005 to perform genomics research in sepsis using a large, well-characterised patient cohort (www.ukccggains.com). Ethics approval was granted by the national ethics committee (approvals 05/MRE00/38, 08/H0505/78) and locally approved for the 34 UK-based participating centres. Recruitment closed in 2018, having enrolled more than 2300 patients, focusing primarily on individuals with sepsis due to community-acquired pneumonia (CAP) or faecal peritonitis (FP). Adult patients (>18 years old) with severe sepsis admitted to ICU in participating hospitals were enrolled by research nurses and written consent was obtained from the patient or legal representative. Sepsis was diagnosed according to Sepsis-2 guidelines (Bone et al., 1992) [**Table 7**]. Clinical status was recorded in an electronic Case Report Form (eCRF) at days 1, 3, 5 and 7 post-ICU admission using severity of illness (APACHE II) and sequential organ failure assessment (SOFA) scores,

microbiology culture results and antibiotic therapy (Knaus et al., 1985, Vincent et al., 1996). Survival status was recorded at 7 days, 28 days and 6 months post-ICU admission.

Clinical Criteria (Sepsis-2)	
▪ Temperature:	>38ºC or <36ºC
▪ Heart rate:	>90 beats/minute or >2 standard deviations above the normal value for age
▪ Respiratory rate:	>20 breaths/minute (tachypnea) or PaCO2 <32 mm Hg (hyperventilation)
▪ White blood cell count:	>12,000/cu mm, <4,000/cu mm, or >10% percent immature neutrophils
▪ Altered mental status	
▪ Evidence of infection	
Exclusion Criteria	
▪ Pregnancy	
▪ Enrolment in an interventional study	
▪ Advance directive to withhold or withdraw life-sustaining treatment or admission for palliative care only	
▪ Pre-existing immune-compromise	
Disease Classification	
Community-Acquired Pneumonia (CAP)	Febrile illness characterised by cough, sputum production, breathlessness, leukocytosis and radiological features of pneumonia acquired prior to, or within 2 days of hospital admission.
Faecal Peritonitis (FP)	Inflammation of the peritoneal membrane secondary to faecal contamination. Diagnosed by laparotomy.

Table 7. The Genomic Advances in Sepsis study inclusion criteria.

Patients were enrolled in GAinS based on the Sepsis-2 diagnostic guidelines as recruitment preceded the publication of the Sepsis-3 guidelines in 2016. Recruitment focused on individuals with sepsis due to either community-acquired pneumonia (CAP) or faecal peritonitis (FP).

2.1.2 Sepsis Samples

Patients were selected for inclusion in the BCR/TCR repertoire study presented in this thesis from enrolled GAinS study participants. To reduce cohort heterogeneity the following selection criteria were used: Caucasian ethnicity, a diagnosis of CAP and availability of three consecutive processed RNA samples (days 1, 3, 5) at a concentration of ≥10ng/μl. Where possible, patients were chosen to overlap with existing 'omics' datasets generated by the Knight Research Group. 'Cohort 1' was comprised of 42 patients sampled at days 1, 3 and 5 (126 RNA samples total), who were selected to investigate longitudinal immune

dynamics in sepsis. These patients well represent the demographics and clinical characteristics of GAinS participants [**Figure 11**]. 'Cohort 2', was chosen to characterise the dysregulated immune response associated with severe disease and address the ascertainment bias caused by inclusion of only patients with three serial time points in Cohort 1. Cohort 2 included 30 patients who died less than 7 days post-ICU admission, using the same criteria as cohort 1. However, individuals were included regardless of the number of available temporal samples to account for longitudinal sampling truncated by death. Of the 72 patients with sepsis, 44 were male (61%) and 28 were female (39%). Significant patient comorbidities are summarised in the appendices [**Table A. 1**].

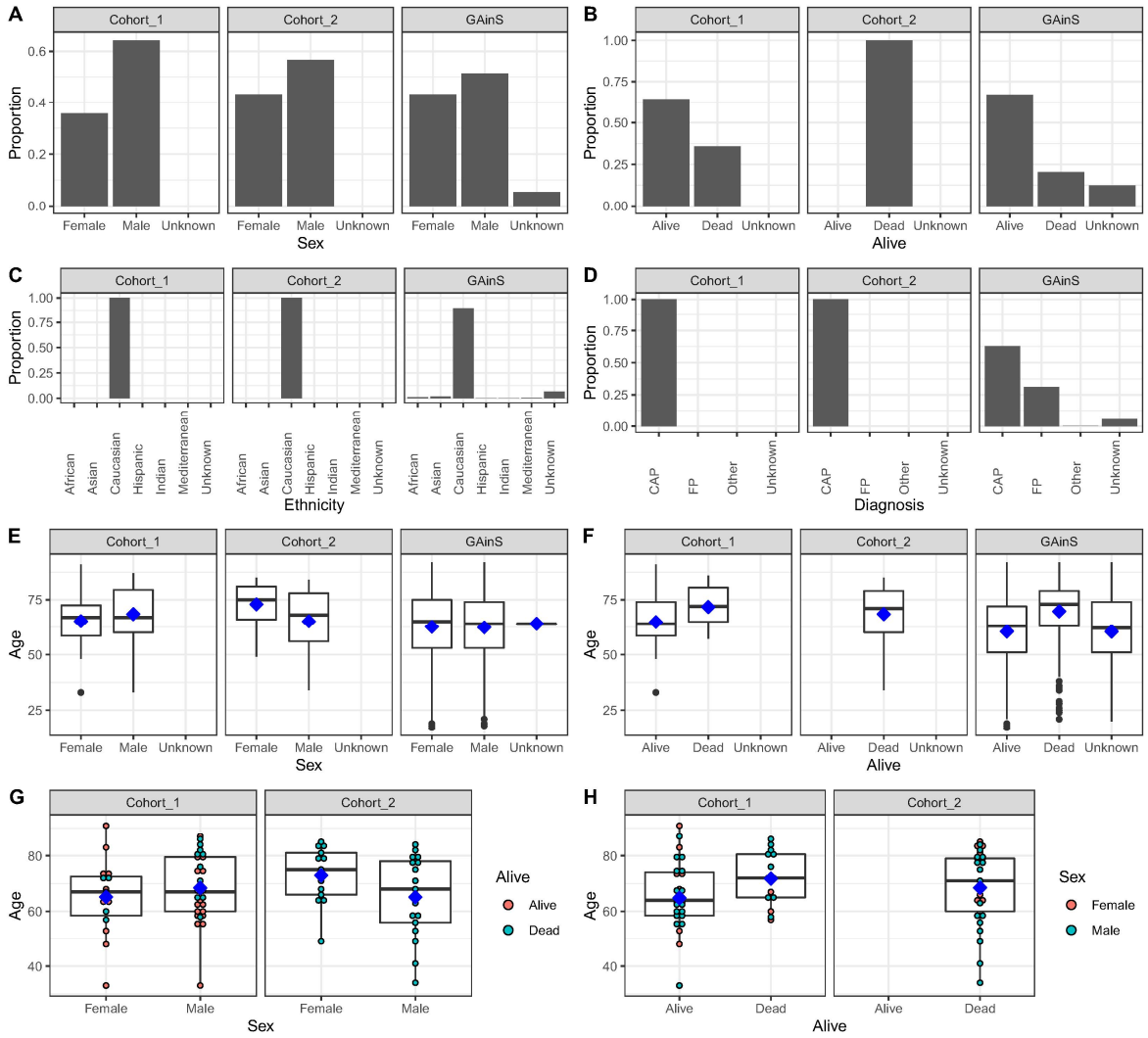


Figure 11. Sepsis patient demographics for adaptive immune repertoire cohorts 1 and 2.

Sepsis patients are compared to all participants recruited to the GAinS study. **A)** The figure shows the presence of a sex bias in sepsis, with male gender potentially having a detrimental effect due to reduced cell-mediated immune response (Angele et al., 2014). **B)** The proportion of 6-month survivors is depicted for each cohort, globally sepsis has an estimated mortality rate of 30-50%. Cohort 2 specifically includes individuals who died within 7 days of ICU admission. **C)** Ethnicity of study participants. **D)** Suspected cause of sepsis among participants. **E, F, G** and **H)** The age of participants grouped by **E** and **D)** sex and **F** and **H**) 6-month survival status. For all plots, the mean is indicated by the blue diamond. Boxplots show the median, first and third quartiles and whiskers are 1.5x the interquartile range.

2.1.3 Healthy Volunteer Recruitment

Healthy volunteers were recruited from the area local to Oxford University, Old Road Campus (Oxford, research ethics committee reference 06/Q1605/55) in 2022. Volunteers were free from active medical conditions or acute infection and had not received a COVID-

19 vaccination in the 3 weeks prior to donation. Where possible, volunteers were sampled on consecutive days 1, 3 and 5 to match the GAinS study. Due to low recall rate, sampling was performed with a minimum of 24 hours between donations, spanning at least 3 days. Age, sex and smoking status was recorded for healthy volunteers. Recruitment timeline and patient demographics are summarised in **Table A. 2** and **Table A. 3** of the appendices. For each sample, 30mls of whole blood was drawn to a total volume of ~100mls across the three time points. Volunteers received no personal benefit for participation, but were reimbursed for travel expenses. Sample handling was matched to the GAinS study to the best of our ability, using the same processing protocol and LeukoLOCK Total RNA Isolation System.

2.1.4 Sample Collection

Sepsis blood samples were drawn by GAinS research nurses and processed at the bedside. Healthy volunteer blood samples were collected by Dr Patrick MacLean and Dr Bo Sun (The University of Oxford) and processed in the laboratory by myself. 10mls of whole blood was collected in EDTA vacutainers and the total blood leucocyte population was isolated using the LeukoLOCK Total RNA Isolation System (Thermo Fisher Scientific) according to the manufacturer's guidelines. In summary, whole blood samples were passed through the LeukoLOCK filter using a vacuum. Filters were washed with 3mls of phosphate buffered saline (PBS) and stabilised with 3mls of RNAlater. Filters were stored at -80°C for long term stability.

2.2 BCR/TCR Multiplex PCR Repertoire Amplification

2.2.1 Preventing Contamination

Samples and reagents were handled in a PCR workstation (AirClean). The PCR workstation and pipettes were decontaminated by 15 minutes of ultraviolet (UV) irradiation and cleaned

with 80% ethanol, DNAZap and RNaseZap (ThermoFisher) prior to use. Filter tips were used to prevent cross-contamination. No post-PCR products were processed inside the PCR workstation.

2.2.2 RNA Extraction

RNA was extracted from thawed LeukoLOCK filters using the LeukoLOCK Total RNA Isolation Protocol (Thermo Fisher Scientific). Cell lysis was performed by passing a guanidine thiocyanate-based solution over the filter using a syringe. The contents of the filter were collected in a 50ml falcon tube and treated with DNase I and Proteinase K to degrade DNA and proteins. RNA was isolated from the treated solution using magnetic bead based extraction. RNA quality and concentration was measured using spectrophotometry (NanoDrop). RNA was stored in 100µl aliquots at -80°C. Healthy volunteer samples were extracted by myself. Sepsis samples were periodically processed by members of the Knight Research Group.

2.2.3 Reverse Transcription

2.2.3.1 Input RNA

Reverse transcription (RT) was performed using the larger volume of either 100ng or 5µl of RNA per sample for BCR IgH repertoire amplification, 150ng or 5µl of RNA for TCRα or β repertoire amplification and 500ng or 5µl of RNA for TCRγ or δ repertoire amplification. A minimum of 5µl of RNA was used to account for lymphopenia in sepsis. RNA was diluted to a total volume of 11µl with molecular grade nuclease-free water (ThermoFisher). 21 samples, one positive control (technical replicate JR-1975-1003) and two reverse transcription (RT) negative controls were used per plate. 1µl of positive control (JR-1975-1003) RNA (~100 ng/µl) was used in each BCR plate and 1.5µl in each TCRα/β/γ/δ plate.

Plate columns 1, 5 and 9 were used to reduce the risk of cross-contamination. RNA-H_2O plates were prepared in advance, centrifuged at 1000xg for 1 minute and stored at -80°C.

2.2.3.2 Primer Annealing

The RNA-H_2O plate was thawed and centrifuged at 1000xg for 1 minute. Meanwhile a master mix containing 1µl of dNTPs (10mM, Promega) and 1µl of RT primer mix (2µM) per sample was prepared. 2µl of master mix was added to each sample, on ice. Plates were incubated at 65°C for 5 minutes in a MiniAMP Plus thermocycler, then chilled on ice for ~60s.

RT primers target the receptor constant region. TCRα and TCRβ RT primers were pooled one-to-one for a combined TCRαβ RT primer mix (final concentration 2µM) enabling simultaneous reverse transcription. TCRγ and TCRδ RT primers were also pooled one-to-one (final concentration 2µM). The BCR RT primer mix (2µM) contained a one-to-one pool of immunoglobulin isotype specific primers. BCR and TCR primer sequences are recorded in **Table A. 4** and **Table A. 5**.

2.2.3.3 cDNA Synthesis

4µl of 5X SuperScriptIV buffer (ThermoFisher), 1µl of dithiothreitol (DTT 0.1M, ThermoFisher), 1µl of RNaseOUT (40U/µl, Invitrogen) and 1µl of Superscript IV (200U/µl, ThermoFisher) per sample were premixed and added to the samples on ice. Samples were mixed by pipetting and the plate was centrifuged at 1500xg for 1 minute. Reverse transcription was performed in the MiniAMP Plus thermocycler using the following RT protocol: 55°C for 50 minutes then 80°C for 10 minutes.

Then, 1µl of RNAseH (2U/µl, Invitrogen) was added to each sample, mixed by pipetting and samples were centrifuged at 1500xg for 1 minute. Plates were incubated at 37°C for 20

minutes in the thermocycler to degrade residual RNA. Incubation proceeded immediately with cDNA clean-up.

2.2.3.4 Bead-Based Purification

36µl of AMPure beads (1.8X, Beckman Coulter) per sample was added to the cDNA and mixed thoroughly by pipetting (x20). The mixture was incubated at room temperature for 10 minutes then transferred to a magnet and settled for 2 minutes. The cleared solution was aspirated and discarded and the plate was removed from the magnet. 40µl of molecular grade nuclease-free water (ThermoFisher) was added to the beads. Beads were re-suspended by mixing thoroughly using pipetting (x20). Plates were transferred back to the magnet and settled for 2 minutes. The supernatant containing purified cDNA was aspirated, transferred to a clean plate and either used immediately for multiplex PCR or stored at -20°C.

2.2.4 Multiplex PCR

2.2.4.1 Primer Design

2.2.4.1.1 BCR

Six IGHV PCR forward primers were designed to target the conserved framework region 1 (FR1) of all known IGHV genes annotated on IMGT as of August 2019. Each set of six forward primers incorporated 1 of 12 unique internal PCR barcodes to enable multiplexing of up to 12 samples. The single CNUS PCR reverse primer was designed to be complementary to a tag on the RT primer, enabling unbiased amplification of immunoglobulin isotypes [**Table 8, Figure 12**]. Primers were validated in Bashford-Rogers et al. (2019). BCR PCR primer sequences are detailed in **Table A. 4**.

Chain	Number of PCR Forward Primers
TCR Alpha	34
TCR Beta	23
TCR Gamma	4
TCR Delta	3 + 6 A/D V gene primers
BCR	6 (x12 barcode sets)

Table 8. The total number of V gene primers for BCR and TCR multiplex PCR reactions.

Primers were designed using the IMGT V gene reference sequences downloaded in 2019.

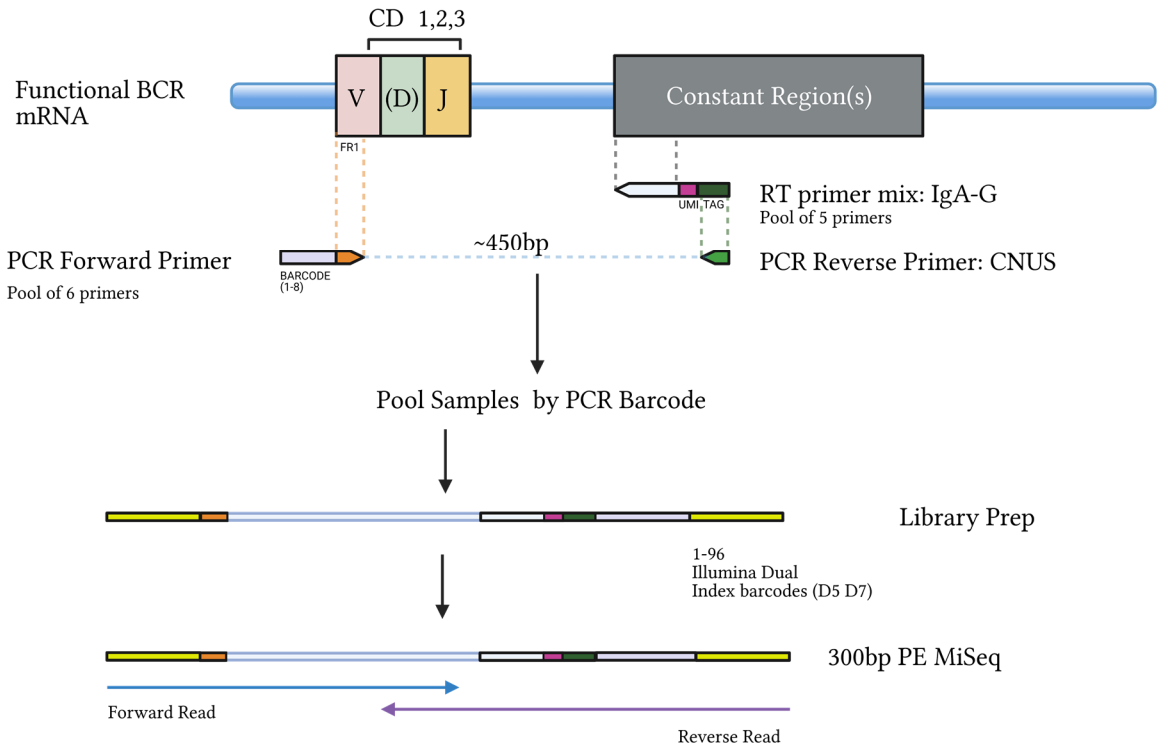


Figure 12. In-house BCR repertoire multiplex PCR amplification protocol schematic.

The multiplex PCR method published and validated in Bashford-Rogers et al. (2019) was optimised as part of this thesis. A pool of reverse transcription primers specific to each immunoglobulin isotype was used to reverse transcribe RNA and incorporate a unique molecular identifier (UMI) into each cDNA molecule. The subsequent step involved amplifying the cDNA using multiplex PCR. The PCR forward primer mix consisted of six primers targeting the framework region 1 (FWR1) of the IGHV gene for all annotated V genes. The forward primers contained an internal PCR barcode for multiplexing of samples. The reverse PCR primer was designed to bind to a tag on the reverse transcription primer, ensuring unbiased amplification of each immunoglobulin isotype. Following amplification, library preparation was performed, and the amplicons were sequenced using 300 base pair paired-end (PE) MiSeq sequencing. The expected size of the amplicons was approximately 450 base pairs in length.

2.2.4.1.2 TCR

For each TCR chain, forward primers were designed to bind all known V genes documented on IMGT (as of August 2019), ideally within a few nucleotides of the 5' region **[Table 8]**. The 'CNUS2_BCX' (TCRα/β) and 'TCR_BCX_DG' (TCRγ/δ) reverse primers bind a tag on the RT primer and include one of 12 unique barcodes (X) to allow multiplexing of up to 12 samples **[Figure 13]**. TCR α and β V gene forward primers were designed by Dr Felicia Tucci (Bashford-Rogers Group, The University of Oxford) and γ and δ V gene primers were designed by myself. Primer sequences are detailed in **Table A. 5**.

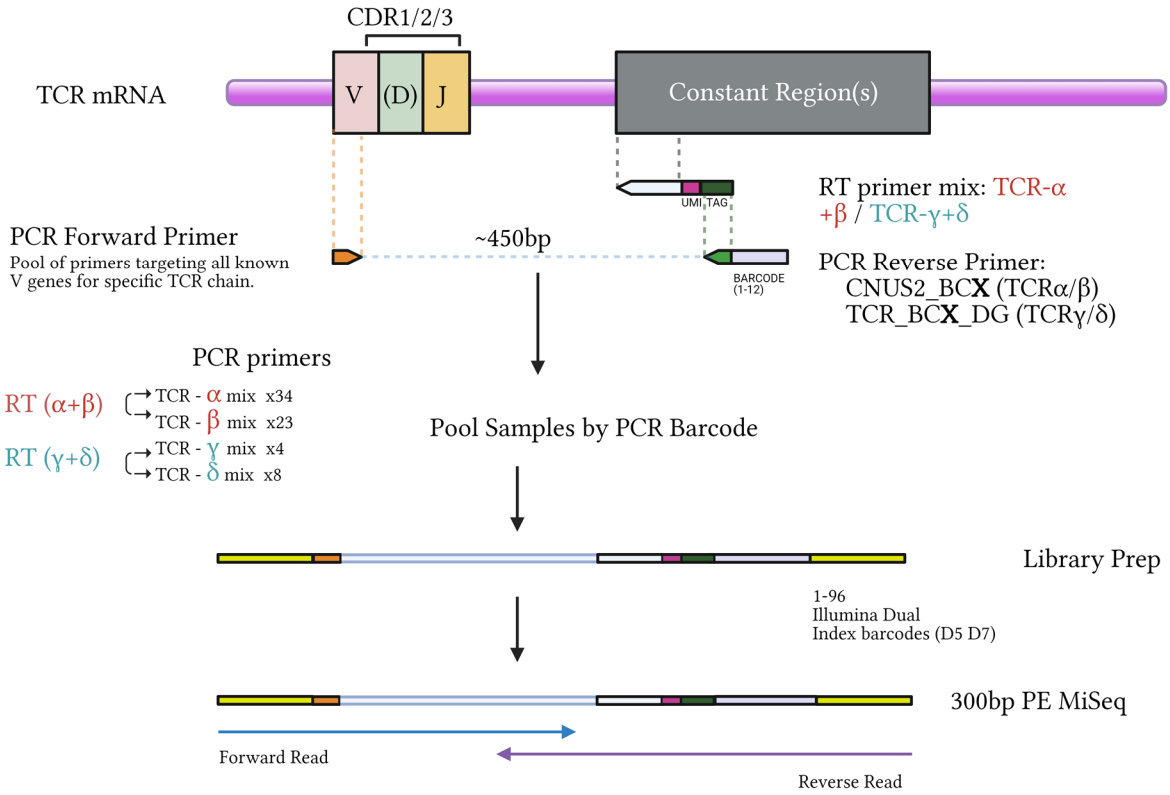


Figure 13. In-house TCR repertoire multiplex PCR amplification protocol schematic.

Reverse transcription was conducted using a pool of either α and β or δ and γ constant region primers. During reverse transcription, a unique molecular identifier (UMI) was incorporated into each cDNA molecule. To enable separate amplification of each TCR chain, the cDNA was divided into two 20µl aliquots. Multiplex PCR was then performed for each chain individually, utilising a mix of forward primers specific to all known V genes and a reverse primer which binds a tag on the reverse transcription primer. To facilitate sample multiplexing, the PCR reverse primer included an internal barcode (1-12). Following amplification, library preparation was conducted, and the amplicons were sequenced using 300 base pair paired-end (PE) MiSeq sequencing. The expected size of the amplicons was approximately 450 base pairs.

2.2.4.1.3 Primer Evaluation

TRG and TRD candidate multiplex PCR and RT primers were designed and evaluated using the web-based Oligo Analysis Tool (Eurofins) and OligoEvaluator (Sigma-Aldrich). Primers were selected based on comparable melting temperatures, a GC content of ~50% and 'none-to-weak' secondary structure. High performance liquid chromatography (HPLC) purified primers were ordered from Sigma-Aldrich.

2.2.4.2 BCR IGH Amplification

25µl of 2X KAPA HiFi HotStart Ready mix (Roche), 1.33µl (10µM) of CNUS reverse primer and 2.4µl of molecular grade nuclease-free water (ThermoFisher) per sample was premixed and centrifuged. The mix was added to 20µl of cDNA on ice. 1.33µl of barcoded forward primer mix (10µM, barcodes 1-8) was added to each sample separately to a total volume of 50µl. In each column of 8 samples, each barcode was used in only one sample. Samples were mixed by pipetting and centrifuged at 1500xg for 1 minute then transferred to the thermocycler for PCR. The following protocol was used: 95°C for 3 minutes, 28 cycles of 98°C for 20s, 62°C for 30s and 72°C for 1 minute and then finally 72°C for 7 minutes [**Figure 12**]. PCR amplicons were stored at -20°C.

2.2.4.3 TCRα/β/δ/γ Amplification

Multiplex PCR was performed separately for each TCR chain. 50µl of 2X KAPA HiFi HotStart Ready mix (Roche) and either 4.5µl (3µM), 3.2µl (4.3µM), 1.3µl (10µM) or 1.3µl (10µM) of TCRα/β/δ/γ forward primer mix respectively, were premixed per sample. Molecular grade nuclease-free water (ThermoFisher) was added to a total volume of 77.3µl per sample. The PCR mixture was centrifuged briefly then added to 20µl of cDNA from the respective reverse transcription reaction on ice. 2.66µl of barcoded reverse primer

'CNUS2_BCX' (10μM, barcodes 1-8, TCRα/β) or 'TCR_BCX_DG' (10μM, barcodes 1-8, TCRγ/δ) was added to each sample separately to a total volume of 100μL. Samples were mixed by pipetting and centrifuged at 1500xg for one minute, then transferred to a thermocycler for PCR, following the protocol detailed in **Table 9.**

Number of Cycles	Temperature	Duration
1 cycle	95 °C	5 min.
5 cycles	98 °C	5 sec.
	72 °C	2 min.
5 cycles	98 °C	5 sec.
	65 °C	10 sec.
	72 °C	2 min.
TCR α/β/γ: **32 cycles**	98 °C	20 sec.
TCR δ: **33 cycles**	**TCR α/β:** 58°C **TCR γ/δ:** 57.5 °C	1 min.
	72 °C	2 min.
Final extension	72 °C	7 min.

Table 9. TCR repertoire PCR protocol.

TCR multiplex PCR cycle numbers and annealing temperatures for each TCR chain.

2.2.4.4 PCR Optimisation

Each chain-specific multiplex PCR reaction was extensively optimised as part of this study, due to the challenges of amplifying RNA from precious sepsis samples where lymphopenia is common. I specifically aimed to reduce the amount of input RNA required, increase amplicon yield and minimise off-target binding. Optimisations and resulting changes to the original Bashford-Rogers et al. (2019) BCR protocol are described in **Table 10.**

Optimisation	Method	Conclusion
PCR annealing temperature	Gradient PCR	Chain specific annealing temperature detailed in the methods. Examples shown in **Figure A. 3.**
RT annealing temperature	Gradient PCR	55ºC was optimal for all BCR/TCR chains. Examples shown in **Figure A. 3.**
cDNA clean-up	o Column-based (QIAquick PCR Purification Kit) o Bead-based (AMPure Beads, Beckman Coulter). o Ran complete protocol.	Column-based clean-up yielded marginally more PCR product, but was not scalable for high-throughput projects. Therefore bead-based clean-up (AMPure) was used. Examples shown in **Figure A. 2.A.**
RNA input quantity	Serial dilution using RNA from healthy and sepsis samples (500ng, 250ng, 150ng, 100ng). Ran complete protocol.	Chain specific RNA input requirement detailed in the methods. Examples shown in **Figure A. 2.C.**
PCR reaction volume	Trialled a total multiplex PCR reaction volume of either 30µl, 50µl or 100µl. Visualised using an agarose gel.	o **BCR:** 50µl reaction produced a good amount of product without requiring a large volume of (costly) reagent. Volume can be increased to 100µl for challenging samples. o **TCR:** 100µl reaction produced a good amount of product.
PCR cycle number	Trialled 25, 28, 30 and 33 PCR cycles. Visualised using an agarose gel.	Chain specific optimal cycle number detailed in the methods and **Table 9.**
cDNA input volume	Trialled 10µl, 15µl and 20µl cDNA input into PCR reaction. Ran complete protocol.	20µl cDNA is optimal for both BCR and TCR. Examples shown in **Figure A. 2.B.**
Paired TCRγδ and TCRαβ PCR reaction	PCR, gel visualisation and trial sequencing.	Combined PCR reactions produced more off-target binding and a lower percentage of reads passed filtering and QC. Therefore separate PCR reactions were used for each chain.
Paired TCRδγ and TCRαβ RT reaction	Trialled 1) separate and 2) combined RT reactions. cDNA was then split in two for chain specific PCR reaction. Ran complete protocol.	For both TCRαβ and TCRγδ similar PCR yield was achieved per chain comparing the combined RT reaction to the separate RT reaction. Therefore RT reactions were combined for each receptor to decrease the total amount of input RNA needed for each sample.
Primer design	Designed two different primer sets for TCRγ and δ, for both RT and PCR. Ran complete protocol.	Primer set 1 failed to amplify TCRδ. Primer set 2 was used. TCRγ primer set 2 increased yield and was used. Examples shown **Figure A. 3.A.**

Table 10. Optimisation of the BCR and TCR multiplex PCR amplification protocol conducted as part of this thesis.

Various experimental methods were tested, and the one that resulted in the largest number of PCR amplicons and the fewest off-target products, as measured by gel electrophoresis, was chosen. Optimisation was first performed on healthy volunteer samples, and then on a small number of sepsis patient samples to ensure the method was robust to low lymphocyte RNA concentration.

2.2.5 Gel Electrophoresis

10µl of PCR product and 4µl of 5X Blue-Juice Gel Loading Buffer (Invitrogen) were mixed and loaded on a 2% agarose gel stained with SYBR Safe. A 100bp ladder (Bioline) was run in parallel to estimate cDNA concentration and amplicon size. Gels were run at 120V for 20-25 minutes and visualised using a ChemiDoc Bio-Rad imager. The expected size of PCR products was 400-500bp in length.

2.2.6 Library Preparation

2.2.6.1 Sample Multiplexing

PCR products were pooled in a one-to-one ratio in autoclaved PCR clean 1.5µl Eppendorfs. Eight samples with unique barcodes 1-8 were pooled into each tube using all of the PCR product from the sample with the lowest DNA concentration. Negative controls and unamplified samples were excluded.

2.2.6.2 Bead-Based Purification

0.8X AMPure beads (Beckman Coulter) were added to each multiplexed pool and mixed thoroughly by pipetting (x20). Tubes were incubated at room temperature for 10 minutes then transferred to a magnet and left to settle for 5 minutes. The cleared solution was aspirated and discarded. Beads were washed twice using 1ml of 80% ethanol (prepared fresh), while on the magnet. Residual ethanol was aspirated and samples were left to air dry, with lids open, for 5-7 minutes. Tubes were removed from the magnet and 55µl of molecular grade nuclease-free water was added to each tube and mixed thoroughly to re-suspend the beads. Tubes were incubated at room temperature for 2 minutes, then transferred to the magnet and left to settle for 2 minutes. The cleared supernatant (55µl) was carefully aspirated, taking care not to dislodge the beads and transferred to new

autoclaved PCR clean 1.5µl Eppendorfs. Pooled samples were stored at -20°C prior to library preparation.

2.2.6.3 Quantification

The concentration of each sample pool was determined using Qubit Fluorometric Quantification (ThermoFisher) according to the manufacturer's protocol. Pools were also visualised and quantified by running 4µl of the product on a 2% agarose gel as in **2.2.5**. 50µl of each pool was input into KAPA Hyper Prep Library Preparation.

2.2.6.4 KAPA Hyper Prep Library Preparation

Sequencing libraries were prepared using the KAPA HyperPrep Kit (Roche) following the manufacturer's protocol and using KAPA dual-indexed adaptors. AMPure beads (Beckman Coulter) were used instead of KAPA beads. For all libraries, 5µl of the 15µM unique adaptor stock was used, without dilution, to account for inaccurate quantification due to primer-dimer and off-target binding. Adaptor ligation was optimal using an interval of 2 hours at room temperature. 4µl of post-ligation product and 4µl of post-amplification product were run side-by-side on a gel as in **2.2.5** to check for successful library preparation, corresponding to ~120bp increase in amplicon length. Libraries were eluted in 25µl of molecular grade nuclease-free water.

2.2.6.5 Library Quantification and Pooling

Library quality was assessed using the 4200 TapeStation (Agilent) high sensitivity DNA assay with 1µl of 1 in 10 library dilution, according to manufacturer's protocol. Library concentration was calculated using the gel from **2.2.6.4**. Libraries were pooled for sequencing into PCR clean 1.5µl Eppendorf's in a one-to-one ratio using a minimum of 5µl per multiplexed pool. Each sequencing pool contained ~11 libraries (corresponding to

approximately 80 multiplex samples) to ensure equal and adequate read coverage during sequencing. Where possible, each pool contained either BCR, TCRα and TCRβ or TCRδ and TCRγ samples. This was done to address variations in PCR efficiency and differing cell proportions in whole blood, both of which impact the quantity of V(D)J amplicons produced by the multiplex PCR reaction. Additionally, this approach aimed to reduce potential sequencing batch effect between the paired chains.

2.2.7 Sequencing

TCR/BCR amplicons were sequenced on a MiSeq (Illumina) using 300bp paired-end read technology (Oxford Genomics Centre, OGC) by project managers Dr Theodosios Kyriakou and Dr Angela Lee. 15% PhiX was used to compensate for low base diversity in libraries and improve cross-talk and phasing calculations. Computational library de-multiplexing was performed by OGC.

2.3 Bulk RNA Sequencing

Samples described in section **2.1.2** and **2.1.3** for which there was no or incomplete temporal bulk RNA sequencing data generated by previous GAinS projects, were submitted for bulk RNA sequencing at the Wellcome Sanger Institute, Cambridge in collaboration with Katie Burnham and Emma Davenport (The University of Cambridge). This cohort will be referred to as 'RNA-seq cohort 2'. Samples were processed using the NEB Automated RNA protocol without globin-depletion. In brief, RNA samples were plated onto two 96-well plates by myself and Dr Eddie Cano-Gamez (Knight Group, The University of Oxford). The smaller volume of either 100ng or 5µl of RNA per sample was diluted to a total volume of 50µl in nuclease free water according to the NEB Automated RNA product sheet, prior to sample submission. cDNA libraries were prepared by the Sanger sequencing facility using the

NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and sequenced using a NovaSeq 6000 (Illumina) on 2 lanes of an S4 flow cell (100bp, paired-end).

RNA-seq cohort 2 is a continuation of a larger dataset previously generated by the GAinS study which includes bulk RNA sequencing for 864 sepsis samples (667 patients). Data is published in Cano-Gamez et al. (2022). The original cohort was processed using the same protocol as described above and was also sequenced at the Wellcome Sanger Institute, Cambridge. Adaptive immune repertoire cohort 1 and cohort 2 individuals described in section **2.1.2,** where all available time points were previously sequenced as part of the 2022 study were not re-sequenced. Instead the existing gene expression data was utilised in this thesis; these individuals comprise 'RNA-seq cohort 1'.

2.4 Computational Analysis

2.4.1 Bulk RNA Sequencing

2.4.1.1 Pre-processing Pipeline and Quality Control

RNA-seq cohort 2 FASTQ files were pre-processed by Katie Burnham, who performed pre-processing for RNA-seq cohort 1, as published in Cano-Gamez et al. (2022). In brief, reads were aligned to the reference genome (GRCh38 v99) using STAR v2.7.3 (ENCODE recommended parameters) and then quantified using FeatureCounts. Counts were normalised and transformed into log2 counts per million plus 1 (log2(CPM+1)) using trimmed mean of M-values (TMM) normalisation in edgeR. Genes that were lowly expressed and did not have ≥10 reads in ≥12 samples (the smallest biologically relevant subgroup comprised of healthy individuals) were excluded.

Six samples were excluded due to low read depth, UK21920120_3, UK19720082_1, UK09010106_1, HV_116_3, UK09320114_5 and UK45430079_1. One sample,

UK55580003_5, was excluded due to low RNA quality (high rRNA percentage and a low gene coverage). A scatter plot of *XIST* (ENSG00000229807) gene expression (expressed only in females) and Y gene usage (found only in males) was used to identify sex mismatch. For all 3 RNA samples from GAUKRV027000 and both samples from UK01310070 the recorded sex (female) did not match the transcriptomic sex (male). For GAUKRV027000 the sex-mismatch was supported by genotyping data, although this was not available for UK01310070. Clinical metadata was updated accordingly.

2.4.1.2 Identification of Sample Mismatch

By comparing matched bulk RNA-seq data with existing genotyping data generated by the Knight Group (where available) using the 'Match BAM to VCF' (MBV) tool from QTLtools, two instances of RNA sample mismatch were identified (Fort et al., 2017). RNA samples UK02870104_5 and GAUKRV025000_3 did not match to their corresponding genotype, whereas the other longitudinal samples from these individuals did. The most likely cause is mislabelling of RNA tubes during the original RNA extraction, and these samples were consequently excluded from both bulk RNA sequencing and repertoire analysis.

2.4.1.3 Batch Effect Assessment

To identify batch effect between RNA-seq cohorts 1 and 2, the log transformed gene expression matrices for both cohorts were merged and the genes common to both datasets (n=19,715) were retained. The 'M3C' R package was used to perform dimensional reduction using uniform manifold approximation and projection (UMAP), t-distributed stochastic neighbour embedding (tSNE) and principal component analysis (PCA) applied to the merged counts matrix [**Figure 14**]. Visualisation of new sample coordinates for the first 2 dimensions shows clear separation of healthy (all sequenced in RNA-seq cohort 2) and sepsis

individuals. Sepsis samples from both cohorts are interspersed and show no evidence of batch effect. Therefore, in downstream analysis the data was treated as one cohort, without a batch covariate.

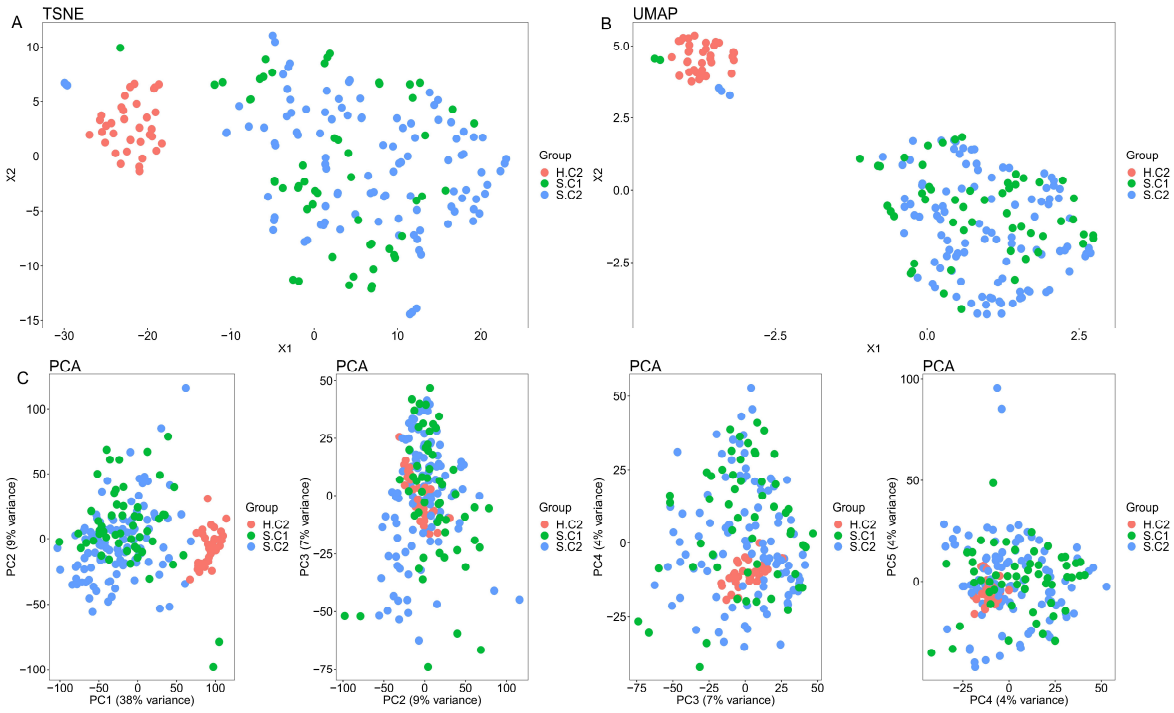


Figure 14. Bulk RNA sequencing batch effect assessment.

Dimensional reduction was performed on log transformed, TMM normalised gene expression matrices, retaining genes present in both datasets (n=19,715), using **A)** t-distributed stochastic neighbour embedding (t-SNE), **B)** uniform manifold approximation and projection (UMAP) and **C)** principal component analysis (PCA). Points are coloured by RNA-seq cohort (C1 and C2) and disease status – health (H) or sepsis (S). Sepsis individuals from both cohorts are well mixed, indicating minimal batch effect.

2.4.1.4 SRS Assignment

Sepsis response signature (SRS1/2/3) and SRSq scores were assigned using the 'SepstratifieR' R package and the Davenport seven gene signature as part of the pre-processing pipeline [**Figure 15**] (Davenport et al., 2016, Cano-Gamez et al., 2022). Across both RNA-seq datasets, as expected, all healthy control longitudinal samples were assigned to SRS3 (n=35) while sepsis longitudinal samples were a mixture of SRS1 (n=71), SRS2 (n=77) and a small number were SRS3 (n=14).

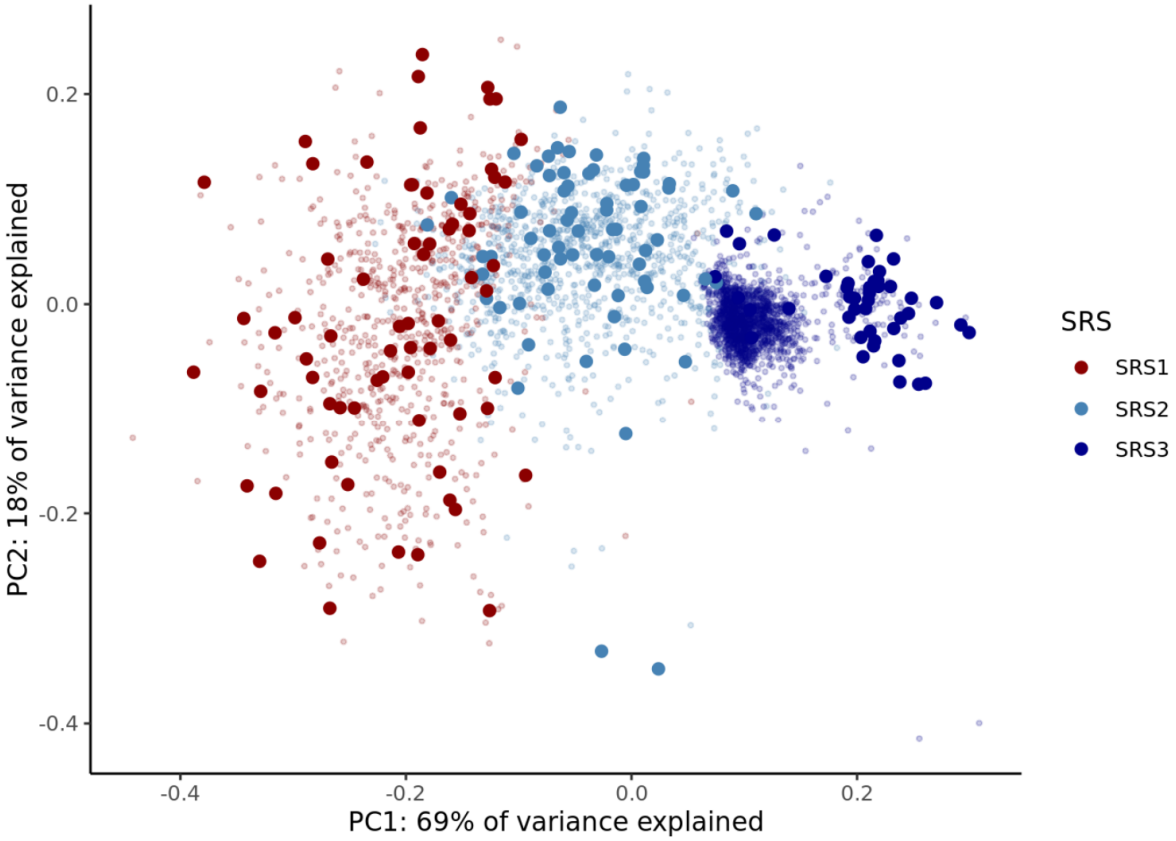


Figure 15. SRS assignment from bulk RNA sequencing gene expression data, using the 'SepstratifieR' R package.

Cohort 2 samples were aligned to the in-built reference and stratified into SRS groups, based on the gene expression of the Davenport 7 gene signature. SRSq score was also calculated. The samples (represented by large points) were projected onto the reference (represented by small points) using PCA to facilitate alignment, assignment, and visualisation. Cohort 1 samples had been assigned an SRS score previously in Cano-Gamez et al. (2022). All healthy individuals were assigned as SRS3.

2.4.1.5 Profiling Cellular Abundance using CIBERSORTx

Blood cellular composition was estimated from the bulk RNA sequencing data using CIBERSORTx (Steen et al., 2020). CIBERSORTx leveraged a sepsis single cell RNA sequencing reference generated by Kwok et al. (2023) as part of the GAinS study. Cell deconvolution was performed by Eddie Cano-Gamez and Hannah Blundell (Knight Group, The University of Oxford).

2.4.2 Pre-processing: Bulk BCR/TCR Repertoire Analysis

2.4.2.1 Summary and Code Availability

Sequencing reads were pre-processed using a custom in-house analysis pipeline described in sections **2.4.2.2-2.4.2.5** and summarised in **Figure 16**. Pre-processing code was developed in Python 2.7 by Rachael Bashford-Rogers (Bashford-Rogers et al., 2016, Bashford-Rogers et al., 2019, Bashford-Rogers et al., 2014). I developed a BASH pipeline manager to automate job submission on the BMRC computing cluster (SLURM workload manager), requiring minimal user input. This allows for the parallel processing of all samples, starting from raw FASTQ files, through to the generation of repertoire summary measures. I also developed and incorporated new R code to generate additional repertoire summary measures, perform integrated quality control (QC) analysis to detect batch effect and sequencing contamination and collate and filter repertoire feature matrices. Code is available from: www.github.com/Bashford-Rogers-lab/Bulk_Repertoire_Analysis.

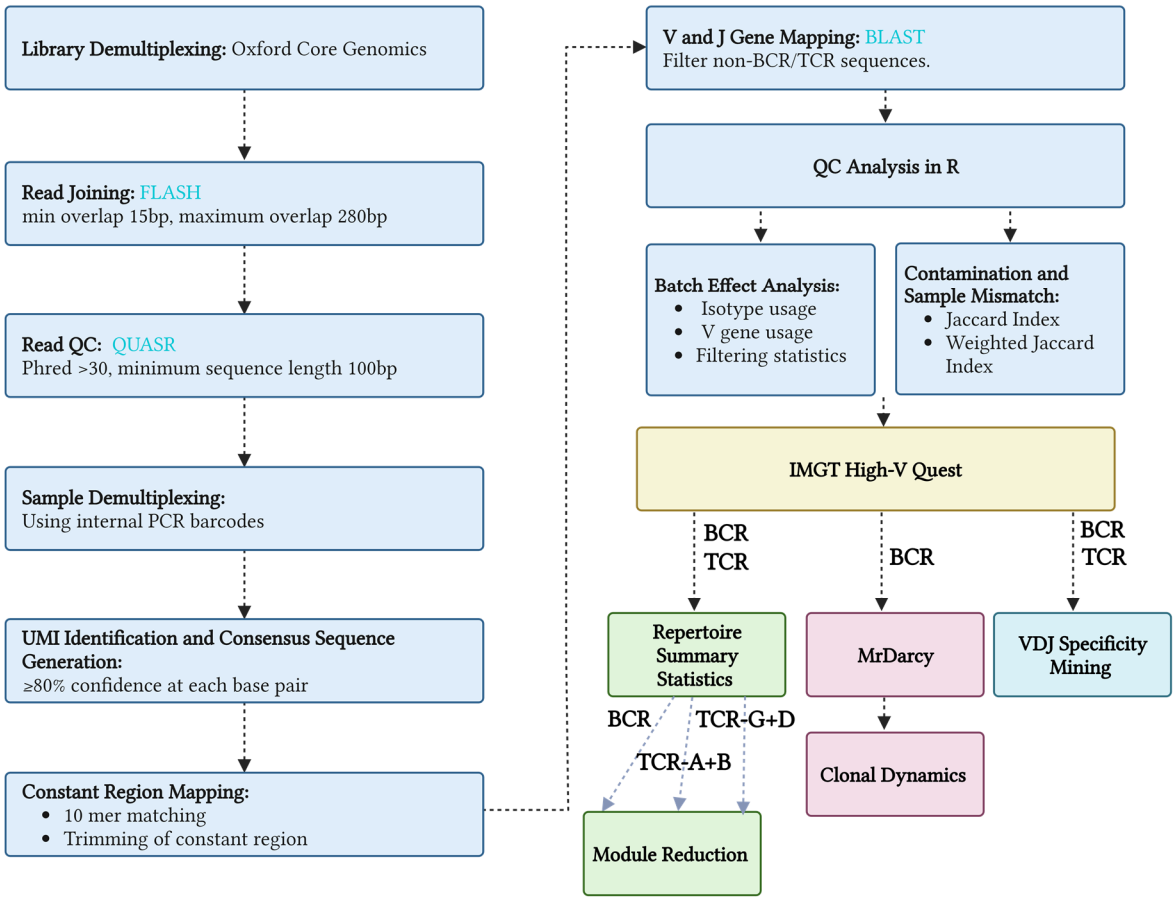


Figure 16. In-house bulk BCR/TCR computational analysis pipeline.

The process begins with pre-processing (depicted in blue) from raw FASTQ files, followed by the manual upload of filtered BCR/TCR sequences to IMGT HighV-QUEST for gold standard annotation. The annotated reads are then input into downstream analysis pipelines, which include the generation of immune repertoire features for dimensional reduction (shown in green), the investigation of clonal dynamics using MRDARCY (shown in red), and the exploration of receptor antigen specificity.

2.4.2.2 Quality Control and Joining Paired Reads

Paired reads from raw sequencing data were joined using FLASH (www.sourceforge.net/projects/flashpage/files/). Minimum and maximum sequence overlap parameters were 15bps and 280bps respectively. QUASR was used to filter joined reads for base quality: a minimum median Phred score of 30 and minimum read length of 100bp was used (www.sourceforge.net/projects/quasr/).

2.4.2.3 Sample De-multiplexing

PCR V gene forward primers and the RT universal primer were identified in each read enabling read orientation from V gene to constant region. The internal PCR barcode was used to de-multiplex filtered reads into sample-specific FASTA files.

2.4.2.4 Generating a Consensus Sequence

The unique molecular identifier (UMI) was identified by screening each read for the conserved bases which constitute the constant portion of the RT primer and flank the UMI. Reads containing identical UMIs were grouped. If there were <5 sequences in the group and $\geq 80\%$ confidence in nucleotide identity at each base pair the group consensus sequence was retained. If there were ≥ 5 reads in a group, sequences were first clustered. The largest cluster was used to generate a consensus sequence using the $\geq 80\%$ confidence in nucleotide identity rule. Reads from groups with $<80\%$ confidence or forming smaller secondary clusters were discarded. Primers were trimmed from each sequence.

2.4.2.5 VDJ-C Mapping

V(D)J and constant region usage was assigned based on sequence similarity to a constant-region gene reference downloaded from IMGT (up-to-date as of August 2019), using 10-mer matching. Constant regions were trimmed, leaving only the V(D)J sequence. Constant gene usage was encoded in the unique sequence identifier for downstream analysis. Sequences were then aligned to an IMGT V/J gene reference using BLAST (www.blast.ncbi.nlm.nih.gov/Blast.cgi). Reads of sufficient similarity to the reference were retained.

2.4.2.6 The Jaccard Index

The Jaccard index (JI) was calculated for all pairwise sample comparisons using **Equation 1**. To account for differences in read depth the JI was calculated on subsamples of the repertoires at a read depth equal to $0.9 \times$ the minimum read depth across the dataset. This was repeated 10,000 times and a mean JI was calculated (Bashford-Rogers et al., 2019). The JI was calculated for both unique and total V(D)Js (weighted index accounting for multiple copies of the same V(D)J due to CSR). The weighted index may improve identification of sample contamination where samples contain large clonal expansions.

$$J(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{intersection\ of\ (unique\ or\ total)\ VDJ\ sequences}{union\ of\ (unique\ or\ total)\ VDJ\ sequences}$$

Equation 1. The Jaccard index was used to quantify V(D)J sequence sharing between samples.

An elevated Jaccard index can arise from library index hopping or contamination as well as convergent antibody response.

2.4.2.7 IMGT HighV-QUEST

Filtered putative V(D)J sequences were uploaded to IMGT HighV-QUEST for gold-standard annotation (https://www.imgt.org/HighV-QUEST/home.action). Each FASTA file was only aligned to the expected BCR or TCR chain. The output files from IMGT were downloaded in CSV format and input into the second half of the pre-processing pipeline for network assembly and repertoire summary statistic generation. Based on the IMGT V(D)J functionality annotations, output files were split into 'productive', 'unproductive' and 'all' reads so that summary metrics could be calculated on each population separately. The focus

of downstream analysis in this thesis is productive reads only. IMGT annotations are described in **Table A. 6**.

2.4.2.8 Network Assembly and Analysis

V(D)J network generation was performed as in Bashford-Rogers et al. (2013). CD-HIT (www.github.com/weizhongli/cdhit) was used to cluster reads with >95% sequence identity. Within each group custom python scripts were used to perform all-against-all pairwise alignments to determine intra-group edges and generate an adjacency matrix. Groups were then compared to determine inter-group edges. Vertexes are proportional to the number of sequencing reads identical to the vertex sequence, excluding differences due to isotype usage. Edges connect vertices that differ by a single nucleotide non-indel difference. Clusters of connected vertices are termed 'clones'. A clone contains a group of related cells with either identical CDR3 and V gene usage or differing by single point mutations to at least one other sequence in the cluster. Each cluster is assumed to represent an evolutionary lineage.

2.4.2.9 Immune Repertoire Feature Generation

Repertoire summary measures for productive reads (**2.4.2.9.2-2.4.2.9.7.1**) were calculated per individual using the IMGT summary files. Summary statistics were concatenated into a repertoire feature matrix for downstream statistical analysis.

2.4.2.9.1 Subsample Depths

All repertoire feature metrics in the following section that are described as using a down-sampling approach to account for differences in read depth, used a user-specified read depth documented in **Table 11,** unless otherwise stated. The selected read depth was a compromise between generating representative repertoire measures (where larger subsample depths are better) and sample dropout due to filtered read counts below those specified.

Subsample depths were chosen with the aim of excluding no more than 10% of samples based on read depth.

Isotype	Sub-Sample Depth
IGH ALL	1500
IGHA2	150
IGHA1	195
IGHGP	20
IGHG1	90
IGHG2	100
IGHG3	20
IGHG4	15
IGHD	165
IGHE	20
IGHM	445
IGHEP2	15
TRAC	300
TRBC	300
TRDC	50
TRGC	50

Table 11. BCR immunoglobulin isotype and TCR chain specific subsample depths.

Samples were excluded from further analysis if their overall read depth was less than 300 for the TRAB repertoire, 50 for the TRGD repertoire and 1500 for the BCR repertoire, which was a compromise between the generation of representative repertoire measures, which is facilitated by increased read depth, and sample drop-out. All measures that used a subsample approach defaulted to the read depths specified in this table, unless otherwise stated.

2.4.2.9.2 Clonal Expansion and Diversification Indices

The output of the network generation algorithm (**2.4.2.8**) was used to calculate different clonal expansion and clonal diversification (a measure of SHM in B cells) metrics using the methods detailed in **Table 12**.

Metric	Summary	Equation
Gini Cluster/Vertex Index	The Gini index was calculated by ordering the cluster/vertex sizes from the smallest to the largest and generating a cumulative frequency distribution where $R = \{r_1, r_2, \ldots, r_n\}$ r_i is the cumulative size of all the largest clusters/vertex until the ith largest cluster/vertex and normalised such that $r_n = 1$. When applied to the vertex size (number of unique RNA molecules per vertex) distribution the Vertex Gini Index measures clonal expansion. When applied to cluster-size distributions (number of unique VDJ sequences per clone) the Cluster Gini Index measures the clonal diversification of a sample. The Gini indices were calculated on an independent repertoire subsample using a fixed depth per isotype and repeated 100 times to calculate a mean Gini Index.	$\boldsymbol{Gini\ index\ (G)} = \sum_{i=1}^{N} \frac{\left(r_i - \left(\frac{i}{N}\right)\right)}{N}$
Rényi Cluster/Vertex Index	Both clonal expansion and clonal diversification indexes were also calculated using the cluster/vertex Rényi Index (α =1) which is the Shannon Entropy. This is calculated from the distribution of the number of unique VDJ region sequences per clone for the Cluster Rényi Index, and the number of unique RNA molecules per vertex for the Vertex Rényi index, within subsampled BCR repertoires at a specified depth. The mean of 100 independent subsamples was calculated.	$\boldsymbol{Renyi\ Index = H_{\alpha}\ (X)} = \frac{\alpha}{1-\alpha} log\left(\sum_{i=1}^{n} p_i^{\alpha}\right)$ $\boldsymbol{Shannon\ Entropy} = -\sum_{i=1}^{n} p_i\ log p_i$
Maximum/Second Maximum and Mean Cluster/Vertex Size	To identify monoclonal or polyclonal expansions the maximum and second maximum cluster and vertex sizes per constant region were identified using the results of the network assembly stage. They are reported as a percentage of the total repertoire. The mean vertex and mean cluster size were also calculated for the whole repertoire.	
D5/D10/D50 Index	The percentage of the repertoire explained by the top 5, 10 and 50 clones was calculated (per constant region) respectively, informing on clonal expansion and overall repertoire diversity.	

Table 12. Clonal expansion and diversification indices.

Methods and descriptions of different measures of clonal expansion and diversification. Adapted from Bashford-Rogers et al. (2013) and Bashford-Rogers et al. (2019).

2.4.2.9.3 CDR Characterisation

Mean CDR3 length was calculated per isotype using the CDR3 regions annotated by IMGT. The CDR3 hydrophobicity index was calculated using the 'hydrophobicity' function from the 'Peptides' R package using the Kyte-Doolittle method (GRAVY index). The GRAVY index is widely used to detect hydrophobic regions in both transmembrane and surface exposed proteins; a positive score indicates that the protein is hydrophobic. Mean CDR2 and CDR3 charge was calculated using the number of R and K amino acid residues for each unique VDJ sequence.

2.4.2.9.4 Distribution of Quantitative Traits

Kurtosis, skewness and the bimodality coefficient were used to characterise the distribution of quantitative features (CDR3 length/hydrophobicity and BCR SHM). Measures were calculated using their respective functions from the 'moments' R package and are summarised in **Table 13**.

Measure	Definition
Kurtosis (γ)	Pearson's measure of kurtosis is a measure of the combined weight of a distributions tails relative to the centre of the distribution. A standard normal distribution has a kurtosis of 3 (adjusted kurtosis 0). Negative kurtosis describes a 'fat-tailed' distribution where extreme outcomes are more common than would be predicted under a normal distribution. Positive kurtosis indicates that values are closer to the mean than expected.
Skewness ($\varkappa$)	Skewness is used to describe the tails of a distribution. A normal distribution has a skewness of 0, a positive skew means the right hand tail is larger than the left hand tail and a negative skew indicates the reverse.
Bimodality Coefficient $\beta = \frac{(\gamma^2 + 1)}{\kappa}$	The bimodality coefficient integrates the size, kurtosis and skew of a distribution, producing a measure which ranges between 0 and 1. Bimodal distributions with light tails or asymmetrical distributions increase the bimodality coefficient. A value >5/9 is often used to indicate a multimodal distribution.

Table 13. Statistical measures used to characterise the distribution of continuous repertoire traits.

2.4.2.9.5 V Gene Usage

V gene usage was calculated using the IMGT HighV-QUEST gene annotations. Where multiple alignments with equal probability were proposed, the first option provided by IMGT was used. The proportion of the repertoire explained by each V gene was calculated using unique V(D)J sequences to account for differences in RNA count per cell. V gene family usage was also calculated per immunoglobulin isotype.

2.4.2.9.6 BCR Specific Measures:

2.4.2.9.6.1 Class-Switching

2.4.2.9.6.1.1 Relative Class-Switch Event Frequency

The relative class-switch event frequency was calculated using the frequency of unique VDJ sequences expressed as two isotypes, indicating that the sequence has likely originated from more than one B cell, one of which has undergone CSR. Overlapping frequencies were measured as the proportion of unique VDJs present as both isotypes IgX and IgY within a random subsample of BCRs at a pre-specified depth. The mean overlap frequency of 1,000 repeats was generated. This method accounts for differences in total read depth between samples but does not account for differences in the relative frequencies of BCRs per isotype.

2.4.2.9.6.1.2 Normalised Class-Switch Event Frequencies

Normalised class-switch event frequencies were calculated by randomly subsampling each isotype to a fixed depth of 50 reads and the proportion of unique VDJ sequences shared between isotype IgX and IgY was calculated. This was repeated 5,000 times and the mean normalised class-switch frequency was calculated.

2.4.2.9.6.2 Isotype Usage

Two measures of isotype usage were calculated: 1) the percentage of reads per isotype, which does not control for differential RNA per cell and can reflect the effect of plasma cells and plasmablasts on the repertoire and 2) the percentage of unique VDJ sequences per isotype, which controls for differential RNA counts per cell, reducing potential biases. Sequences were also grouped by class-switched, IGHM/D un-mutated and IGHM/D mutated categories.

2.4.2.9.6.3 Somatic Hypermutation

The mean SHM was calculated on unique VDJ sequences to account for biases due to differing RNA counts per cell. Several different measures were calculated including: mean mutation rate for different VDJ regions (full length VDJ, CDR, CDR3, FRW, FRW3, and the CDR:FRW ratio), the mean rate of silent and non-silent mutations, the non-silent:silent ratio and the percentage of un-mutated sequences per isotype. The distribution of total VDJ mutations for all sequences within the dataset was plotted using a histogram [**Figure A. 1**]. Sequences were classified according to the peaks and troughs on the distribution, determined by visual inspection, and these classifications were used to calculate the percentage of reads with no, low, high or very high mutation [**Table 14**].

VDJ SHM Classification	Number of Mutations
Un-mutated	$x \leq 3$
Low mutation	$3 < x \leq 11$
High mutation	$11 < x \leq 75$
Very high mutation	$75 < x$

Table 14. The classification of BCR sequences according to the total number of mutations.

Thresholds were chosen based on the distribution of BCR SHM across all reads from all samples, shown in **Figure A. 1.**

2.4.2.9.6.4 Autoreactivity

The percentage of un-mutated IGHV4-34 AVY and NHS motifs were calculated per isotype, as germline AVY and NHS motifs have been linked to B cell auto reactivity. Mutations in these motifs due to SHM abrogate self-reactivity (Schickel et al., 2017).

2.4.2.9.6.5 V Gene Replacement Frequency

V gene replacement frequency was calculated as a percentage of the total number of unique BCRs, by counting the number of stem sequences associated with more than one IGHV gene with >95% difference in sequence identity and using different 5' junctional regions. Frequencies were calculated on a subsample of BCR sequences and repeated 50 times to generate a mean frequency (per isotype). Stem regions are defined as N-IgHD-N-IgHJ regions starting 3bp downstream of the IGHV gene boundary. Sequences with joining regions (N-IgHD-N-IgHJ) shorter than 8 nucleotides were excluded from this analysis due to the potential for germline-encoded receptors. Although identical rearrangements can happen by chance, it is assumed that this background rate is constant across patients.

2.4.2.9.7 TCR Specific Measures

2.4.2.9.7.1 Invariant T Cell Detection

TCRα sequences were classified according to V and J gene usage as either standard T cell, possible mucosal associated invariant T cell (MAIT), possible invariant NKT cell (iNKT) and possible 'germline-encoded, mycoyl-reactive' (GEM) T cell. The proportions of each type were calculated per sample **[Table 15].** However, in the absence of paired TCRβ chains, the T cell origin of these sequences cannot be confirmed.

T cell Type	Invariant TRAV Genes	Reference
MAIT	TRAV1-2^{+}TRAJ33^{+} or TRAV1-2^{+}TRAJ12^{+} or TRAV1-2^{+}TRAJ20^{+}	Koay et al. (2019)
iNKT	TRAV23^{+}TRAJ18^{+} or TRAV10^{+}TRAJ18^{+}	de Greef et al. (2020), Kitaura et al. (2016)
GEM	TRAV1-2^{+}TRAJ9^{+}	Kronenberg and Zajonc (2013)

Table 15. Classification of TCR sequences according to invariant TCRα V and J gene usage.

2.4.3 Comparing Isotype usage between Bulk RNA Sequencing and Targeted BCR Repertoire Amplification

To quantify the usage of immunoglobulin isotype '*IGHX*' in targeted BCR repertoire sequencing, the percentage of unique (or total) BCRs belonging to *IGHX* in relation to all unique (or total) BCR reads was calculated (**2.4.2.9.6.2**) [**Equation 2**]. However, interpreting this measure can be challenging, as changes in the absolute expression of other isotypes, despite stable expression of immunoglobulin isotype *IGHX* across time, results in a change in the proportion of *IGHX* and the illusion of a change in the absolute expression of *IGHX*. A more accurate interpretation of this measure is the relative usage of *IGHX* compared to all other isotypes.

$$Targeted\ Isotype\ Usage\ (\%) = \frac{Number\ of\ unique\ (or\ total\ reads)\ of\ isotype\ X}{\sum unique\ (or\ total)\ reads\ for\ all\ isotypes} \times 100$$

Equation 2. Isotype usage estimated using targeted BCR repertoire sequencing.

Isotype usage can be quantified more accurately using the counts per million (CPM) gene expression data obtained from bulk RNA-seq. Normalisation of gene expression accounts for individual library size, and each isotype is quantified relative to the total library containing over 19,000 genes. Consequently, changes in the absolute expression of isotype *IGHY* have minimal impact on the normalised expression of isotype *IGHX*. To investigate the reproducibility of isotype usage across methods, the TMM-normalised counts from bulk

RNA sequencing were transformed into a comparable measure similar to **Equation 2.** This was achieved by calculating the percentage of normalised counts for each isotype in relation to the total normalised counts for all immunoglobulin isotypes [**Equation 3**].

$$Bulk\ Isotype\ Usage\ (\%) = \frac{TMM\ normalised\ gene\ counts\ for\ isotype\ X}{\sum TMM\ normalised\ gene\ counts\ for\ all\ isotypes} \times 100$$

Equation 3. A comparable measure of isotype usage was calculated using the TMM normalised isotype gene counts from bulk RNA sequencing.

To account for non-independence of samples, repeated measures correlation (RMCORR) (**2.5.1**) was used to calculate the correlation between the normalised isotype usages obtained from both experimental methods. Strong and statistically significant positive correlations were observed for most isotypes, except for IGHE and IGHG4, which have low frequencies in peripheral blood and are captured relatively infrequently [**Figure 17**]. In contrast, the correlation between the proportion of the BCR repertoire explained by each isotype, as generated from the BCR repertoire processing pipeline, and the CPM isotype gene expression was not significant for IGHM, and showed weak correlations across all other isotypes. This can be explained by biased expansion of class-switched isotypes relative to non-class-switched isotypes [**Figure 18**].

Therefore, I can conclude that normalised isotype usage was captured consistently regardless of experimental approach. However, accurately quantifying 'absolute' changes in BCR isotype expression using the targeted BCR amplification method is not possible. The common practice of normalising immunoglobulin isotype counts as a percentage of all BCR reads in BCR repertoire analysis should be approached with caution, and researchers need

to be aware of the limitations associated with this approach when interpreting results (Zhang et al., 2022, Bashford-Rogers et al., 2019, Galson et al., 2020).

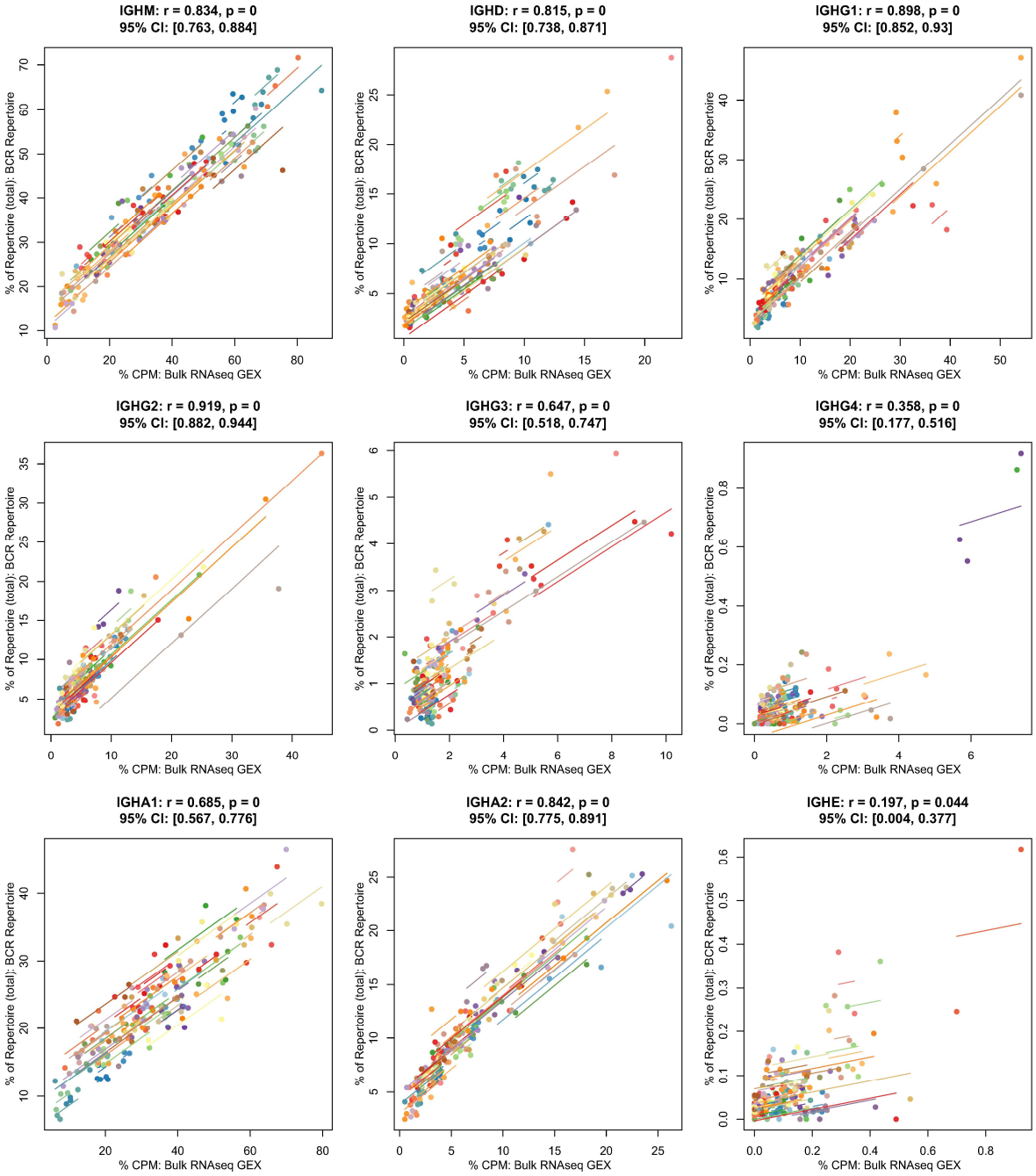


Figure 17. Good correlation between the percentages of the repertoire comprised of each isotype, as output from the targeted BCR repertoire analysis pipeline (y axis) versus the normalised bulk RNA sequencing (x axis) isotype measure.

RMCORR was used to estimate the correlation coefficient (R_{rm}) accounting for non-independence of samples and the coefficient and corresponding p values are reported in the title for each plot, along with the 95% confidence intervals. Each participant's data and corresponding common regression slope are shown in a different colour.

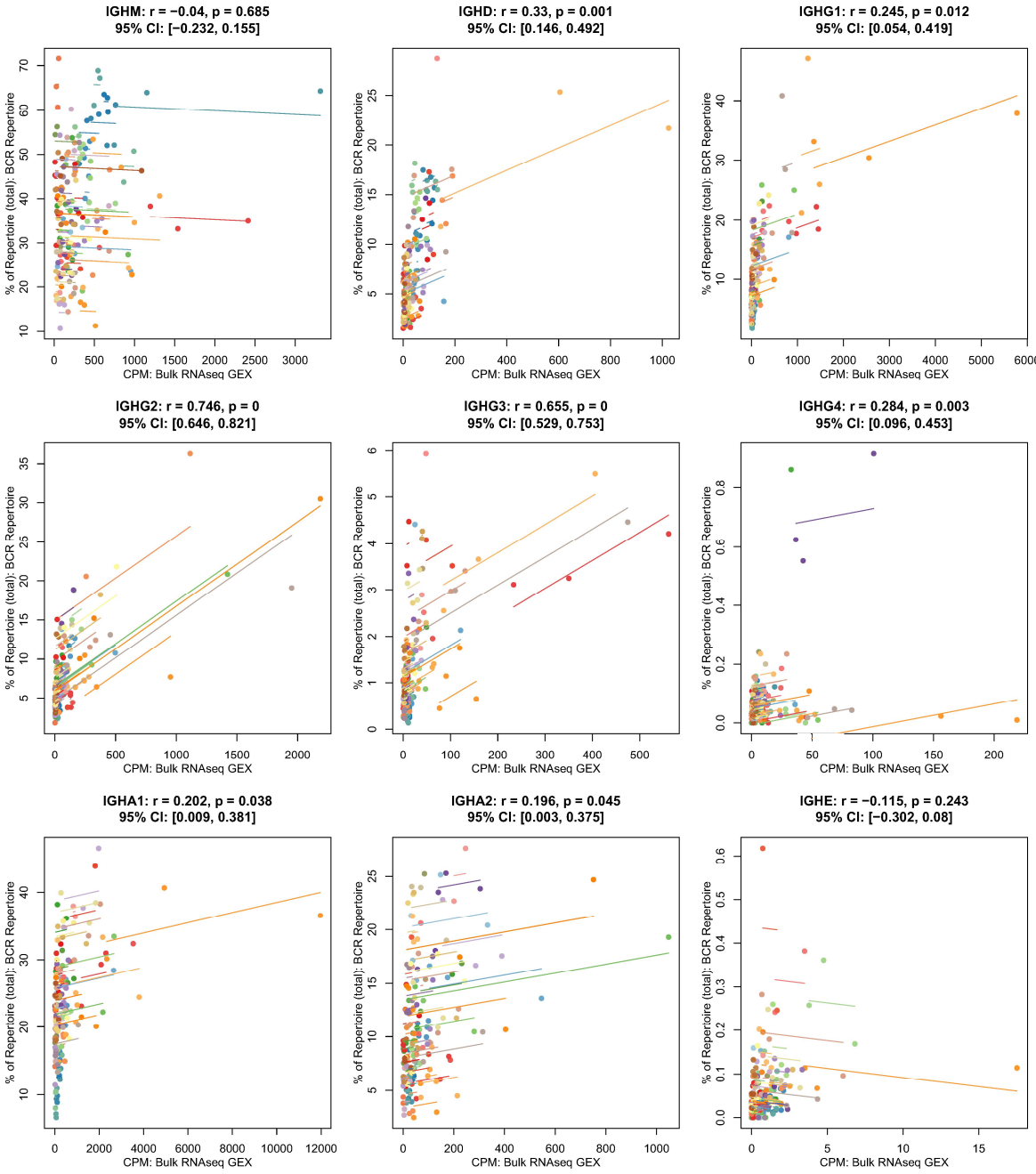


Figure 18. Poor correlation between the percentages of the repertoire comprised of each isotype, as output from the targeted BCR repertoire analysis pipeline (y axis) versus the corresponding immunoglobulin isotype gene expression from bulk RNA sequencing data (x axis).

RMCORR was used to estimate the correlation coefficient (R_{rm}) accounting for non-independence of samples and the coefficient and corresponding p values are reported in the title for each plot, along with the 95% confidence intervals. Each participant's data and corresponding common regression slope are shown in a different colour.

2.5 Statistical Analysis

2.5.1 Repeated Measures Correlation

To account for non-independence of samples due to longitudinal sampling, repeated measures correlation (RMCORR) was used to calculate the correlation between two continuous variables of interest where, for non-paired data, typically Pearson Correlation Coefficient (PCC) or Spearman's Rank would be used. For detailed methodology, refer to Bakdash and Marusich (2017). In brief, RMCORR estimates the common regression slope shared between individuals, similar to a null multilevel model (varying intercept and a common slope for each individual as visualised in **Figure 17** and **Figure 18**). The correlation coefficient ranges from -1 to 1. Absolute values closer to 1 indicate a strong relationship while a coefficient close to 0 indicates little to no association. Typically, RMCORR has increased statistical power as samples are neither averaged or aggregated to account for non-independence, prior to calculation of the RMCORR coefficient (R_{rm}) (Bakdash and Marusich, 2017). RMCORR was calculated using the 'rmcorr' R package.

2.5.2 Linear Mixed Effects Models

2.5.2.1 Accounting for Non-independence of Samples using Random Effect Terms

To identify significant associations between fixed effects such as disease status (categorical) and time point (continuous) and dependent variables such as module score (continuous), whilst correcting for clinical covariates (*e.g.* age) and non-independence of samples, linear mixed effects models (LMEs) were used, utilising the 'lme4' R package (Bates et al., 2015). Models were fitted using maximum likelihood estimation. A random effect term specified as '+ (1|Barcode)' where barcode is the patient identifier, was used to account for non-

independence of samples. This corresponds to a random intercept for each group of patient identifiers (a set of longitudinal samples) to account for individual-specific variation *e.g.* each patient may have a different baseline. It is possible to specify a more complex model with both random intercepts and random slopes. However, given the relatively few number of repeated observations per individual (<3) this led to over fitting and the models would not converge. The lmer4 package does not directly calculate p values due to the complexity of estimating them in the presence of random effects. Therefore, to estimate the significance of model terms, the 'lmerTest' package which is an extension of lme4 was used. P values were estimated using Satterthwaite's degrees of freedom method (Kuznetsova et al., 2017).

2.5.2.2 Interaction Terms and Simple Effects Tests

In multiple different analyses the effect of both time and a categorical grouping variable such as disease status (sepsis versus health or sepsis mortality groups) on a continuous outcome variable was investigated using linear mixed effects models. In these cases, an interaction term was also included to determine whether the effect of time differed significantly between groups. In the case of a significant interaction, main effect terms such as time cannot be interpreted directly as they depend on the levels of interacting variables. Therefore, follow-up simple effects tests were used, breaking the interaction into separate models at each level of the categorical variables and performing pairwise comparisons across all factor levels. This allows for a more detailed understanding of the interaction term and is a common approach in the scientific literature (Garofalo et al., 2022). For example, when investigating the difference in continuous outcome variable A, between health and disease, the five simple effects tests models are:

1) Three models comparing the difference in A between health and sepsis at each level

of time (day 1, day 3 and day 5) using simple linear models, adjusting for covariates.

2) Two models investigating whether A changed significantly across time in health or sepsis using LME models and a random effect term to account for non-independence of samples, adjusting for covariates.

However, one of the limitations of this approach is that it represents an over-simplification of the interaction effect and fails to consider how values are modulated by all levels of the factors simultaneously (interaction is a global effect). A second drawback is the need for multiple testing correction, as the cumulative probability of finding one of the comparisons significant by chance increases dramatically with the number of tests performed. However, most methods for multiple testing correction such as the Bonferroni and Benjamini-Hochberg methods make the assumption that tests are independent, which is not the case here (Lee and Lee, 2018). Hence, both the raw and Benjamini-Hochberg adjusted p values are reported.

It is generally accepted that descriptive interpretation of the estimated means, having removed variance attributed to other main effects or covariates, is an effective method to represent and report interaction effects. This allows for both the answering of specific hypotheses and data exploration, without incurring a multiple testing burden (Garofalo et al., 2022). Therefore, interaction models were also visualised using partial residual plots, from the 'interactions' R package. These plot partial residuals after controlling for the estimated variance attributed to other factors included in the input model (covariates and the random effect term). Both visual investigation and simple effects tests can be used in conjunction to understand the interaction effect.

2.5.2.3 Mortality Analysis

To account for missing data due to longitudinal sampling truncated by death, the guidelines detailed in Kurland et al. (2009) were followed. A fully conditional pattern mixture linear mixed effects model was employed, meaning the dependent variable was described separately for groups defined by survival time (essentially stratifying by time of individual dropout). A random effect term was also included to account for sample non-independence. Patients were grouped into three mortality groups: 7-day non-survivors (early-death, ED), 6-month non-survivors (late-death, LD) and 6-month survivors. However, a limitation of this modelling approach is the use of future survival information to predict the earlier group response, hence these models cannot be used to predict an individual's trajectory based on baseline information.

2.5.2.4 Variance Inflation Factors

Variance Inflation Factors (VIF) were used to test for multicollinearity between predictor variables in statistical models using the 'car' R package. Predictors that are significantly correlated can inflate standard errors and p values, leading to a lack of statistical significance. Multicollinearity may require that some predictors are removed from the model. A threshold of VIF$>$5 was used as indication of significant multicollinearity (Marcoulides and Raykov, 2019).

2.5.3 Multiple Testing Correction

To account for multiple testing, the Benjamini-Hochberg (BH) adjustment was used, which is a simple method that controls the False Discovery Rate (FDR). The false discovery rate is the expected number of false positives out of all tests (type I error) and is used where a low number of type I errors are acceptable (Lee and Lee, 2018). The BH method is less

conservative than the Bonferroni method which controls type II error and can often fail to detect real differences, especially where there are many tests or the test statistics are positively correlated. As noted in **2.5.2.2,** both methods assume tests are independent and where this assumption is violated both raw and BH adjusted p values are reported.

2.6 B Cell Clonal Dynamics with MRDARCY

2.6.1 Method Summary

MRDARCY (MRD Assessment and Retrieval Code in pYthon) was used to identify persistent B cell clones across longitudinal samples, as in Bashford-Rogers et al. (2016). In brief B cell clone detection was performed by clustering VDJ sequences using CD-HIT (0.85 similarity) allowing up to 8bp mismatches. Clusters with identical CDR3s (>5bp in length) and V and J gene family identities were merged. The percentage of the repertoire explained by each clone was calculated for each sample, to account for differences in sequencing depth.

2.6.2 Clonal Dynamics

In order to characterise clonal dynamics across time, clones were classified as either 'increasing', 'decreasing' or exhibiting an 'insignificant change', using the percentage change in repertoire (x) between the first and last available time point ($\Delta x_{overall}$). The percentage change in clone size was calculated for each time point transition available, to understand how sample dropout affects the distribution of Δx. This was important as some individuals had only day 1 and day 3 samples, while others had all three. $\Delta x_{overall}$ is expected to be greater when first and last time points are separated by a larger time period [**Table 16**]. To account for stochasticity in clone size with time, due to repertoire resampling, amplification and sequencing, a density plot of $\Delta x_{1:3}$, $\Delta x_{3:5}$, $\Delta x_{1:5}$, and $\Delta x_{overall}$ was plotted using every clone within the dataset. The mean ($\bar{x}$) and standard deviation (y) was

calculated for each distribution. For all distributions $\bar{x} \rightarrow 0$ and $y_{1:3} \approx y_{1:5} \approx y_{3:5}$. As the mean tended to zero, clones where $|\Delta x_{overall}| < y$ where y was the standard deviation for the transition used to calculate $\Delta x_{overall}$ for that sample (*e.g.* day 1 and 3, day 1 and 5 or day 3 and 5) were classified as 'insignificant change'. Clones where $\Delta x_{overall} > \mathrm{y}$ were classified as 'increase' and for clones where $\Delta x_{overall} < -1 \times \mathrm{y}$ the movement was 'decrease'.

Notation	Calculation	Definition
$\Delta x_{1:3}$	$x_3 - x_1$	Change in percentage of the repertoire for clone X between day 1 and 3.
$\Delta x_{3:5}$	$x_5 - x_3$	Change in percentage of the repertoire for clone X between day 3 and 5.
$\Delta x_{1:5}$	$x_5 - x_1$	Change in percentage of the repertoire for clone X between day 1 and 5.
$\Delta x_{overall}$	$x_{last\ time\ point} - x_{first\ timepoint}$	Change in percentage of the repertoire for clone X between the first and last available time point.
$y_{1:3}$		Standard deviation for the change in percentage of the repertoire for all clones between day 1 and 3.
$y_{3:5}$		Standard deviation for the change in percentage of the repertoire for all clones between day 3 and 5.
$y_{1:5}$		Standard deviation for the change in percentage of the repertoire for all clones between day 1 and 5.

Table 16. Classification of B cell clonal dynamics using the change in percentage of the repertoire between the first and last available time points.

2.6.3 Clonal Overlap

To investigate clonal sharing across time points the Jaccard index (JI) was calculated for the top 10 clones at the first and last time points, which represent the largest percentage of the repertoire for each individual [**Equation 4**]. The JI was also calculated using all clones from an individual by down-sampling each time point to 300 clones, calculating the clonal overlap between time points using the JI and repeating 10,000 times to calculate a mean clonal overlap.

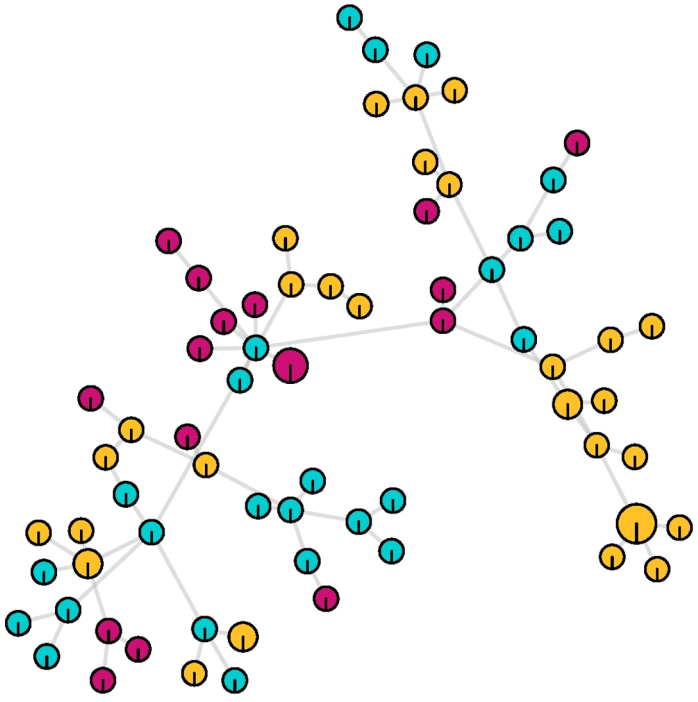
$$J(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{intersection\ of\ clones}{union\ of\ clones}$$

Equation 4. Clonal overlap measured using the Jaccard index.

2.6.4 BCR Longitudinal Clone Network Generation

BCR networks were generated for each participant separately and on individual BCR clones containing more than 10 unique VDJ sequences spanning more than 1 time point. BCR sequences from individual longitudinal clones were aligned using the 'ape' R package. Provided the aligned sequence length was >80 base pairs, the hamming distance was calculated pairwise for all aligned sequences. Identical unique VDJ sequences which were found at more than 1 time point were collapsed into one node, where node size was proportional to the number of unique copies of that sequence irrespective of isotype. Neighbourhood-joining tree estimation was used to construct an undirected phylogenetic tree which was optimised using maximum likelihood methods implemented in the 'phangorn' R package. For the same BCR clone an undirected minimum spanning tree (MST) was also constructed. A MST connects all vertices together without cycles and uses the minimum total edge weight. For each clone both node-level centrality and network-level centralisation measures (degree and eigenvector) were calculated to investigate B cell clone structure and evolution using the 'igraph' R package. For degree centralisation self-loops were not allowed, therefore the theoretical maximum is a star graph where degree centralisation equals 1. Thus, degree centralisation is independent of the number of vertices. This code was developed in collaboration with Dr Stephen-John Sammut (The Institute of Cancer Research).

Chapter 3. The Development of Computational Methods to Analyse the BCR and TCR Repertoire



3 The Development of Computational Methods to Analyse the BCR and TCR Repertoire

This chapter introduces a novel computational method for dimensional reduction of adaptive immune repertoire data, utilising the principles of weighted gene correlation network analysis, termed repertoire (r) WGCNA. This addresses an unmet need in the field, due to the high-dimensionality and complexity of BCR and TCR repertoire data. IgH repertoire data generated for a cohort of sepsis patients and healthy controls will be used in this chapter as a case study. The performance of rWGCNA is then compared an alternative machine learning (ML) approach, Sparse Decomposition of Arrays (SDA).

3 The Development of Computational Methods to Analyse the BCR and TCR Repertoire 123

3.1 Introduction 123

3.2 Chapter 3 Aims 128

3.3 Cohort Summary 128

3.4 Data Quality Control Analysis 130

3.5 Method Development: Dimensional Reduction of Repertoire Feature Matrices 142

3.6 Discussion 177

3.7 Conclusion 185

***B* Appendices: The Development of Computational Methods to Analyse the BCR and TCR Repertoire 390**

3.1 Introduction

In this chapter, I present the computational analysis conducted on a BCR IgH repertoire sequencing dataset, generated for a cohort of sepsis patients and age-matched healthy controls as part of this thesis. Firstly, I discuss the comprehensive data quality control (QC) analyses, which serve as guidelines for future research. Next, I summarise the methods currently used in the scientific literature to analyse complex high-dimensional adaptive immune repertoire data.

The majority of this chapter is dedicated to the development of a novel computational approach for (longitudinal) adaptive immune repertoire analysis. Specifically, I have adapted and optimised the principles of weighted gene correlation network analysis (WGCNA) to perform BCR and TCR repertoire feature dimensional reduction (**3.5.2**) (Langfelder and Horvath, 2008). The complexity of the adaptive immune repertoire, and the potential for subtle perturbations in the context of disease that may not achieve statistical significance, can make it challenging to identify pathophysiology and incur a significant multiple-testing burden. Therefore, the development of dimensional reduction techniques, particularly those that are able to improve understanding of the biological mechanisms underpinning changes to the adaptive immune repertoire, is important to the field. The custom analysis pipeline, repertoire WGCNA (rWGCNA), presented in this chapter outputs a small number of modules that capture co-regulated features of the BCR repertoire, which inform on shared biology. The biological significance of these modules will be discussed in **Chapter 4**. Finally, I present the advantages and limitations of rWGCNA in contrast to an alternative dimensional reduction method, Sparse Decomposition of Arrays (SDA) (section **3.5.3**) (Hore et al., 2016).

3.1.1 Immune Repertoire Sequencing Analysis

High-throughput adaptive immune repertoire sequencing generates large, complicated datasets, which require specialist analysis pipelines, distinct from those used for conventional RNA-seq analysis. This is because of the complexity and repetitiveness of BCR and TCR loci and the V(D)J rearrangement process. During V(D)J recombination biologically meaningful nucleotide insertions and deletions occur at the junctions between gene segments, causing variation from germline references. In the case of the BCR repertoire, point mutations from SHM add additional complexity (Nouri and Kleinstein, 2020).

To facilitate downstream analysis, accurate annotation of the different regions of the receptor sequence corresponding to the conserved framework and variable complementarity determining regions and V, D and J segment identity is critical (Rosati et al., 2017, Yaari and Kleinstein, 2015). In recent years, numerous computational tools have been developed for immune repertoire analysis (both BCR and TCR). These can be categorised based on three functions: 1) pre-processing and read annotation software, 2) post-processing tools for calculating repertoire summary statistics (outlined in **Table 17**), and 3) representation machine learning techniques for encoding receptor sequences. **Table 18** provides a summary of commonly used software packages.

Currently, there is no universally accepted analysis pipeline or protocol for repertoire-specific QC analysis. Many research groups rely on custom code, often in conjunction with their own experimental protocols (Bashford-Rogers et al., 2019, Bashford-Rogers et al., 2016, Galson et al., 2020). However, IMGT has emerged as the 'gold-standard' annotation platform and reference database, and comprehensive start-to-finish analysis platforms such as Immcantation are gaining popularity (Galson et al., 2020, Stephenson et al., 2021, Waltari et al., 2019, Mathew et al., 2021).

To facilitate sharing of data, scientists are encouraged to use the 'minimum information adaptive immune receptor repertoire (MiAIRR) metadata standards' for the submission of sequencing datasets. These standards were agreed by the AIRR community to guide the publication, curation, and sharing of AIRR-seq data and metadata (Vander Heiden et al., 2018). Consequently, new analysis tools are recommended to use input data in AIRR standard format. Many existing tools incorporate this feature by default. For example, IMGT HighV-QUEST generates summary files in either CSV or AIRR format, compatible with downstream tools including VDJTools, bcRep, and ChangeO [**Table 18**].

Measure	Methods	Biological Relevance	Reference
Clonality	Sequences are clustered based on sequence similarity, forming groups (clones) derived from a common ancestor. In B cells, clonal members may have non-identical VDJ sequences due to somatic hypermutation (SHM), allowing quantification of intra-clonal diversification. Analysis is often restricted to heavy chains to have sufficient diversity to accurately distinguish clones.	**Within an individual**: identification of persistent clones or those spanning different biological compartments. Relevance in cancer immunology and autoimmunity. **Between individuals:** detection of convergent responses, is important for understanding population-level immune responses to infectious diseases and vaccines.	Bashford-Rogers et al. (2016), Hershberg et al. (2014), Nielsen et al. (2020b), Jiang et al. (2020)
Diversity	Diversity has been typically measured using three main approaches: species richness, Shannon Entropy, and the Gini-Simpson index, each capturing different aspects of diversity. These methods have been adapted from ecological studies.	Changes in diversity in relation to clinical status and disease outcome is an active area of research. Loss of diversity may limit response to antigens.	Shifrut et al. (2013), Bashford-Rogers et al. (2019), Mina et al. (2019), Baum et al. (2012), Hou et al. (2016)
V(D)J, C gene usage	Quantification of the frequency and distribution of specific V, (D), J and constant region gene combinations within the repertoire.	Antigen recognition, autoimmunity, and antibody effector function.	Schickel et al. (2017), Bashford-Rogers et al. (2019), Zhu et al. (2011)
CDR3 length distribution	The average length distribution of the complementary determining region 3 (CDR3).	Antigen recognition.	Wang et al. (2017), Rock et al. (1994), Hou et al. (2019a), Fang et al. (2022)
Mutation frequency	The degree of SHM for BCRs is determined by comparing the observed sequence to the closest germline reference. SHM is often calculated per framework and CDR region.	An increase in non-synonymous mutation rate is used as evidence of antigen-driven positive selection. BCR SHM in the context of B cell lymphoma may promote survival.	Xu-Monette et al. (2019), Odabashian et al. (2020), He et al. (2022), Agathangelidis et al. (2022)

Table 17. A compilation of commonly used summary measures for characterising adaptive immune repertoires.

Function	Tool	Application	Reference
Pre-processing and annotation: **Goal:** process raw sequencing reads into error-corrected, annotated VDJ sequences.	**IMGT HighV-QUEST** Note: widely accepted as gold-standard annotation tool.	Annotation of VDJ sequences including V-region mutation, germline V/D/J assignment and the location of framework and CDR regions.	**Alamyar et al. (2012)**
	IgBLAST	Germline V/D/J assignment, delineation of framework and CDR regions.	**Ye et al. (2013)**
	MIGEC	Sequence error correction using unique molecular identifiers (UMI), mapping of V, D, J segments and extraction of CDR3 regions.	**Shugay et al. (2014)**
	MiXCR	Universal framework from raw sequences to quantitated clonotypes including error correction and SHM quantification.	**Bolotin et al. (2015)**
	pRESTO	Processing reads from raw sequences to error-corrected sorted and annotated sequence sets.	**Vander Heiden et al. (2014)**
Post-Processing **Goal:** generate summary measures describing adaptive immune repertoires and understand population structure.	**VDJTools** Input: IMGT HighV-QUEST, MiGEC, IgBlast	A software suite that provides a wide range of immune repertoire post-analysis tasks such as repertoire overlap, diversity analysis and segment usage statistics.	**Shugay et al. (2015)**
	bcRep Input: IMGT HighV-QUEST	Platform for BCR repertoire analysis including clonotype assignment, diversity, gene usage and principle component analysis.	**Bischof and Ibrahim (2016)**
	ChangeO Input: IMGT HighV-QUEST	Suit of utilities for clonal and VDJ assignment, diversity analysis and sequence chemical properties.	**Gupta et al. (2015)**
	VDJviz	Quality control of repertoire sequencing data, as well as analysis of key characteristics such as clonality and VDJ recombination patterns.	**Bagaev et al. (2016)**
VDJ Representation **Goal:** encode receptor sequences to infer antigen binding/cellular origin.	**IGoR**	Likelihood of recombination and SHM scenarios. Quantitatively characterises the probability of receptor generation from cDNA and gDNA.	**Marcou et al. (2018)**
	AntiBERTa	Contextualised representation of antibody sequences (embeddings).	**Leem et al. (2022)**
	immuneML	Machine learning platform for adaptive immune repertoire data.	**Pavlović et al. (2021)**
Start-to-Finish	**Immcantation**	Framework analytical ecosystem, incorporating python and R packages for pre-processing, population structure determination and repertoire analysis such as **Change-O, Alakazam** and **pRESTO.**	

Table 18. Specialised computational tools for analysis of adaptive immune repertoire sequencing data.

3.2 Chapter 3 Aims

The primary objectives of this chapter are:

1. To provide a comprehensive overview of BCR repertoire QC analyses including:
    a. The identification of batch effect.
    b. Quantification of library index hopping and contamination, which can lead to the false-positive detection of public BCR sequences.
2. To develop and optimise a computational pipeline based on weighted gene correlation network analysis (WGCNA) for the dimensional reduction of longitudinal adaptive immune repertoire data, using in-house sepsis BCR IgH sequencing data as a case study. Important aims include:
    a. Identifying the optimal method for imputing missing repertoire data.
    b. Understanding and mitigating the impact of unbalanced paired design on co-expression networks in longitudinal studies.
3. To perform dimensional reduction of sepsis BCR IgH data using an alternative machine learning approach, Sparse Decomposition of Arrays (SDA).
4. To discuss the benefits and limitations of rWGCNA in comparison to SDA.

3.3 Cohort Summary

The BCR repertoire cohort analysed in the following sections is comprised of 205 samples from 72 individuals with sepsis due to community-acquired pneumonia (CAP) and 12 age-matched healthy volunteers [**Figure 19**]. The sepsis samples were collected from ICU patients enrolled in the GAinS study between 2005 and 2018, using Sepsis-2 diagnostic criteria. The total blood leukocyte population was isolated from whole blood samples, drawn on the first, third and fifth day post-admission to a UK intensive care unit (ICU) where available, using the LeukoLOCK Total RNA Isolation System (ThermoFisher). RNA

extractions were performed periodically by members of the Knight Group (The University of Oxford). Patient clinical data was recorded using an electronic case report form (eCRF) and downloaded from the Knight Group LabKey server using the GAinS x1.0.0 'Hazy Hare' release. Healthy donors were also sampled at days 1, 3 and 5 and processed using the same protocol, as part of this thesis.

BCR IgH RNA was reverse transcribed and then amplified using an in-house multiplex PCR reaction. Samples were library prepped and sequenced in batches throughout 2021-2022 using 300bp paired-end MiSeq sequencing. To identify batch effects, a technical replicate (JR-1975-1003) was re-amplified and sequenced with each sequencing submission. A detailed experimental timeline is provided in Appendix **Table B. 1**.

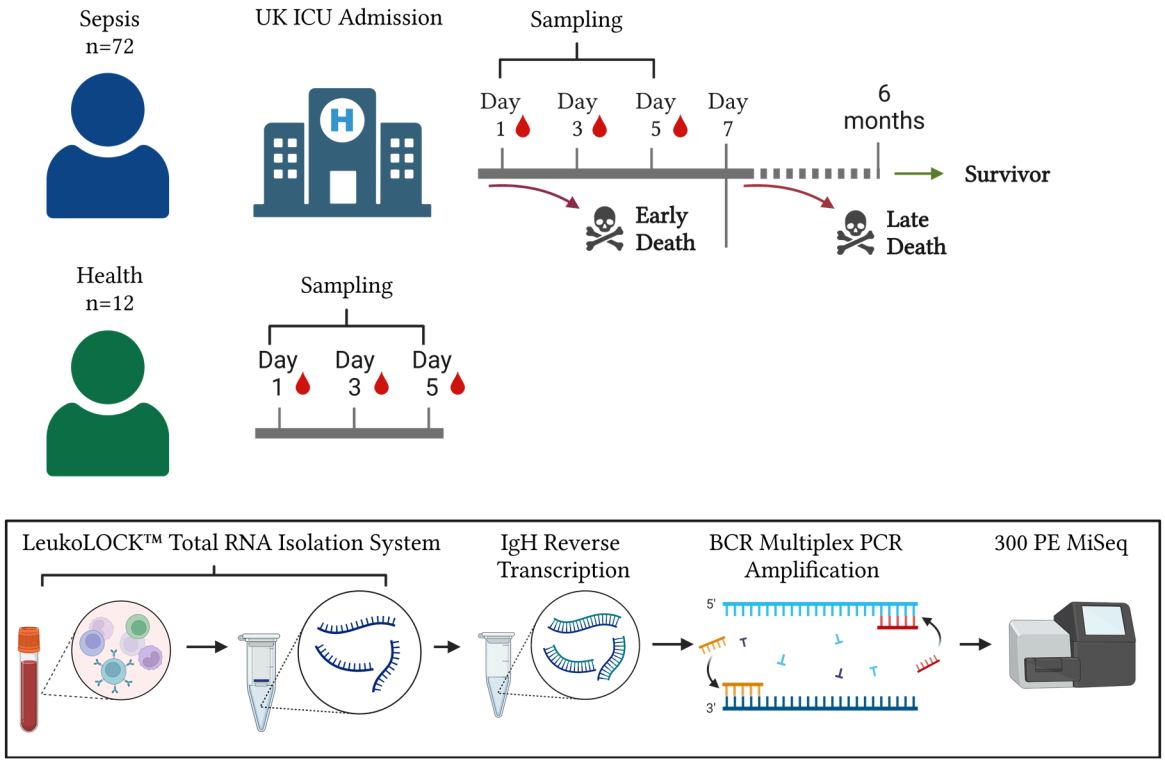


Figure 19. The BCR repertoire in sepsis study schematic.

Sepsis patients were recruited from 34 UK Intensive Care Units (ICU) using Sepsis-2 diagnostic criteria. Blood samples were drawn on days 1, 3 and 5 post-ICU admission. Patients were categorised based on their 7-day and 6-month survival status. Age-matched healthy donors, who were free from active infection or disease and had not received a COVID-19 vaccine in the preceding 3 weeks, were included in the study. RNA was reverse transcribed and the BCR repertoire was amplified using multiplex PCR, prior to paired-end sequencing.

3.4 Data Quality Control Analysis

Data pre-processing was carried out concurrently on all samples, using the custom computational pre-processing pipeline described in **Chapter 2. Materials and Methods** (Bashford-Rogers et al., 2014). In the following section (**3.4**) I summarise the results of the integrated QC analyses, which was developed in R code. QC analysis was conducted on filtered VDJ sequences prior to gold-standard IgH annotation using IMGT.

3.4.1 Batch Effect Assessment

3.4.1.1 Sequencing Depth

The number of unique BCR sequences and the percentage of reads which passed filtering was compared across sequencing lanes and libraries to identify batch effect **[Figure 20, Table 19].** The mean number of unique VDJ sequences post-filtering exceeded 1500 in all libraries. Differences in read depth between libraries are likely due to deviations from one-to-one pooling during library preparation. Accurate DNA quantification was challenging due to off-target binding and primer-dimer contamination. This is reflected in the relatively low percentage of reads which passed filtering (~50-60%). However, libraries performed comparably across sequencing lanes, and the filtering pass-rate was similar to the 60% previously reported by Bashford-Rogers et al. (2014) for PCR amplification and MiSeq sequencing. For future studies, I recommend the use of double-sided size selection rather than single-sided size selection during library preparation, to increase the yield by removing primer-dimers and off-target amplicons (Quail et al., 2009).

At this stage, 11 samples were excluded due to a unique productive VDJ read depth of less than 1500. Repertoire summary measures calculated on samples with low read depth are often unrepresentative of the total repertoire. A further six samples from individuals

UK17730209 and UK59580028 were excluded due to significant comorbidities, metastatic tumour and leukaemia respectively. During pre-processing, substantial clonal expansion was detected for the latter.

There was some evidence that disease status influenced BCR read depth; samples from healthy controls yielded a greater mean number of unique VDJs post-filtering compared to sepsis samples. This is biologically plausible considering the lymphopenia observed in sepsis. However, sepsis samples had been stored, in some cases, at -80°C for over 10 years, which may have led to RNA degradation, exaggerating this difference.

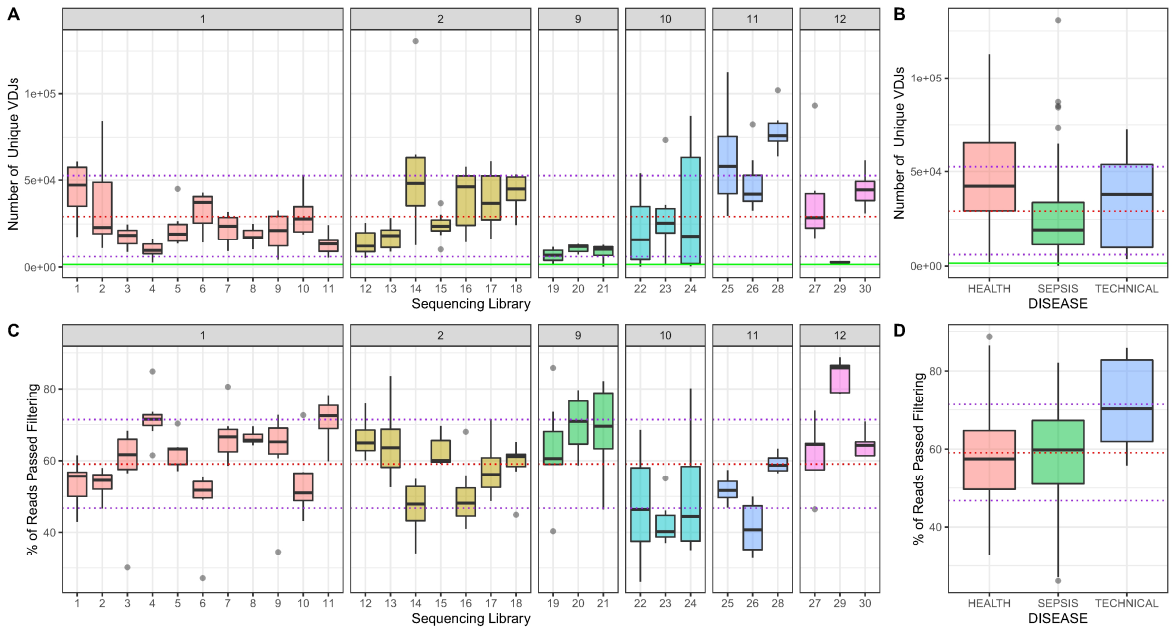


Figure 20. QC summary: BCR repertoire sequencing depth post-filtering, for all samples.

A,B) The number of unique BCRs which passed pre-processing grouped by **(A)** library and sequencing lane and **(B)** disease status. **C,D)** Percentage of BCRs which passed pre-processing grouped as in **A** and **B**. Reads were excluded if they did not map to known IGHV or IGHJ gene segments and/or had low quality (Phred score <30). The red line represents the mean value for each measure across the BCR dataset, while the purple lines indicate the mean ± 1 standard deviation. Samples were excluded if the number of productive unique VDJ sequences was less than 1500 (indicated by horizontal green line). The median, first and third quartile are shown and the whiskers represent 1.5x the interquartile range (IQR).

	No. Unique VDJs Post Filtering			% Reads Passed Filtering		
Lane	**Mean**	**SD**	**NSD**	**Mean**	**SD**	**NSD**
1	24,469.74	15,191.18	0.62	60.72	10.64	0.18
2	32,927.82	21,926.41	0.67	58.26	9.77	0.17
9	8,699.87	3,771.72	0.43	67.16	11.40	0.17
10	27,830.68	27,108.61	0.97	46.84	13.57	0.29
11	62,723.43	23,684.49	0.38	50.46	8.64	0.17
12	29,834.47	24,929.71	0.84	68.86	11.69	0.17

Table 19. QC summary statistics: BCR repertoire sequencing depth post-filtering, for all samples.

The table presents summary statistics including the mean, standard deviation (SD), and normalised standard deviation (NSD=SD/mean) for the number of unique BCR VDJ sequences post filtering and the percentage of reads that passed filtering across all samples. Samples on sequencing lanes 10 and 12 exhibit the highest variation among the dataset. Statistics are reported to two decimal places.

3.4.1.2 IGHC, IGHV and IGHJ Gene Usage

PCR amplification bias can introduce batch effects into BCR sequencing data and may arise when the ratio of multiplexed primers differs due to human and/or pipette error. To identify amplification bias, the percentage of each immunoglobulin isotype was compared across the ten technical replicates included in this study. Constant region usage was similar across lanes, with the exception of the technical included on lane nine [**Figure 21, Table B. 2**]. This replicate had substantially fewer filtered reads, 3,498, compared to a mean of 35,770 across all technical replicates. Low read depth limits the capture of rarer immunoglobulin isotypes, possibly explaining the difference in isotype usage. Replicates on sequencing lanes nine and ten were amplified and sequenced during the same timeframe (**Table B. 1**). As the replicate included on lane ten performed similarly to those on the other lanes, is unlikely that lane nine was affected by significant PCR batch effect. However, the possibility of a sequencing batch effect cannot be disregarded.

IGHJ gene usage was highly reproducible across technical replicates [**Figure 22.A, Table B. 3**]. By comparison, IGHV gene family usage was less reproducible, particularly for the

most prevalent IGHV genes. However, the relative order of IGHV genes when ranked from high to low usage was usually consistent across technical replicates [**Figure 22.B, Table B. 3**]. Absolute differences in IGHV gene usage may be attributed to either batch effect or inaccuracies in IGHV gene annotation during pre-processing. At this stage a relatively relaxed mapping threshold was used with the aim of identifying potential VDJ sequences for downstream analysis and excluding non-VDJ sequences. However, to achieve gold-standard VDJ annotation, all samples were subsequently annotation using the web-based tool IMGT HighV-QUEST.

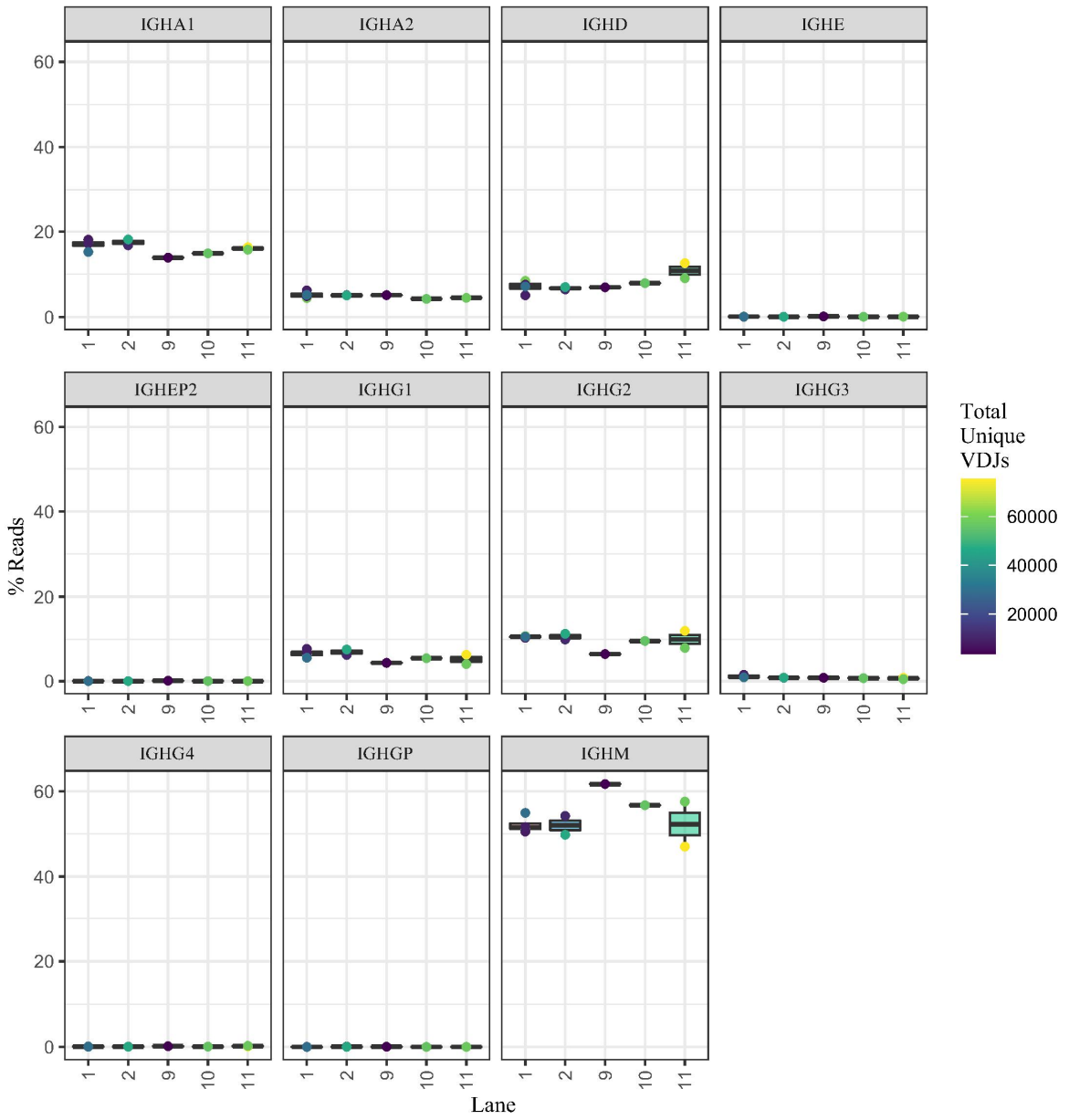


Figure 21. QC summary: immunoglobulin isotype usage for the 10 technical replicates (healthy donor JR-1975-1003) included in the BCR repertoire cohort.

The percentage of total BCR reads for each isotype, per technical replicate, post-filtering. The data points are coloured by the total number of unique VDJ sequences post-filtering across all isotypes for the sample. The median, first and third quartile are shown, while the whiskers represent 1.5 times the interquartile range (IQR). Generally, technical replicate immunoglobulin isotype usage shows little batch effect with the exception of the technical included on sequencing lane 9, which has a comparatively low read depth.

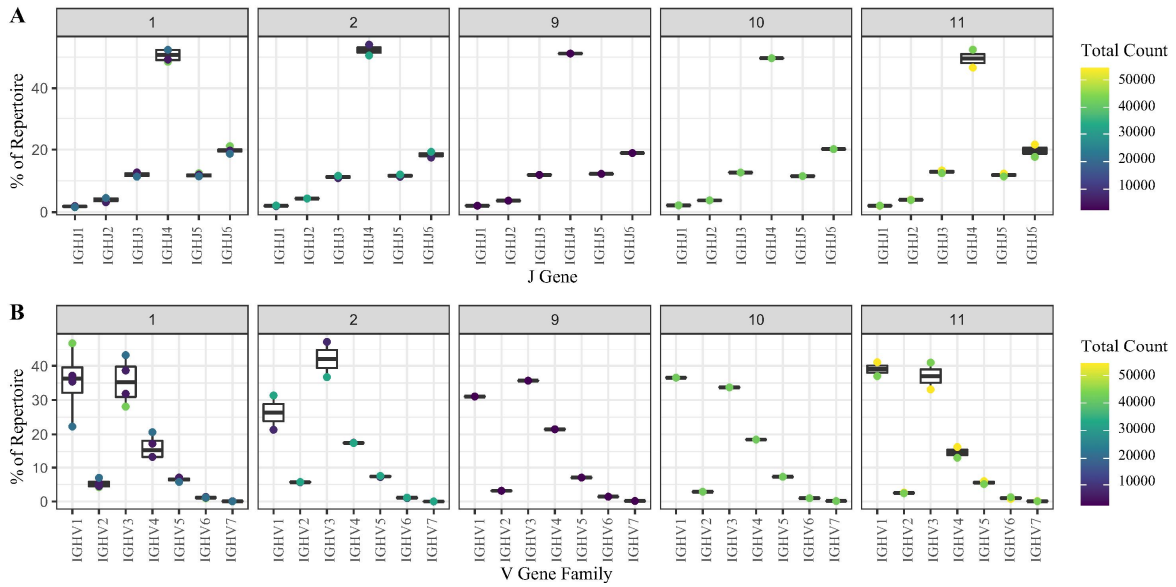


Figure 22. QC Summary: IGHV/J gene usage for the 10 technical replicates (healthy donor JR-1975-1003) included in the BCR repertoire cohort.

The percentage of total unique IgH VDJ reads for each **A)** IGHJ gene and **B)** IGHV gene family per technical replicate, across sequencing lanes. The data points are coloured by the number of unique VDJ sequences for the sample. The median, first and third quartile are shown, and the whiskers represent 1.5 times the interquartile range (IQR). Generally, technical replicate IGHJ gene usage shows little batch effect, although IGHV gene usage is more variable.

3.4.2 Primer Validation

To ensure all IGHV genes were successfully amplified, the total count for each IGHV gene across all samples was calculated. The panel of detected IGHV genes was compared to the IGHV gene reference downloaded from IMGT. All IGHV genes, except for IGHV3-16 (which had an open reading frame) and the pseudogenes IGHV3-29, IGHV3-30-2, IGHV3-30-33, IGHV3-30-42, and IGHV3-32, were detected in the BCR repertoire data [**Figure 23**]. In the context of IMGT, an open read frame (ORF) annotation signifies a coding region with alterations in splicing sites, recombination signals and regulatory elements or changes to conserved amino acids leading to incorrect protein folding [**Table A. 6**]. Pseudogenes contain stop-codons or frame-shift mutations. Therefore, the absence of these genes in the dataset is not concerning. Good read depth was observed across all IGHV gene families;

indicating successful capture of IGHV gene diversity using the in-house multiplex PCR reaction.

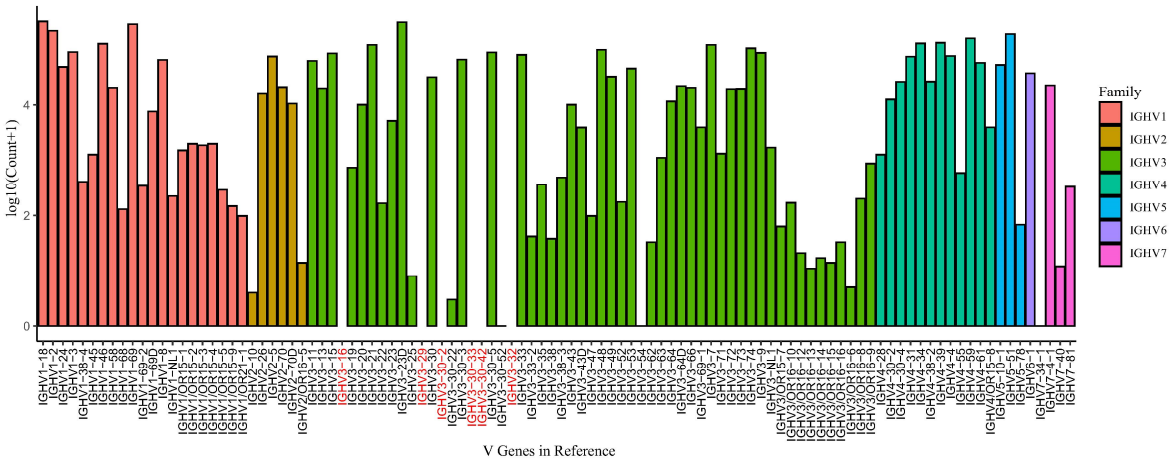


Figure 23. QC summary: total IGHV counts in the BCR repertoire data.

The total IGHV gene count across all samples is plotted. Genes annotated in red on the x-axis are present in the IMGT reference file, but had zero counts in the sepsis BCR cohort and were investigated further, to ensure they were not absent due to experimental error. Individual IGHV genes are coloured according to V gene family. All IGHV gene families are well represented in the data suggesting that the missing IGHV genes reflect population prevalence and not technical error or issues with primer binding.

3.4.3 Index Hopping and Sample Contamination

In multiplexed NGS sequencing libraries, the main cause of incorrect sample read assignment is index hopping. Index hopping is the exchange of the i5 or i7 library index, typically due to free-floating, unligated adaptors that can anneal to cDNA fragments prior to sequencing causing extension of the library fragment with an oligo containing the wrong index (Farouni et al., 2020) [**Figure 24**]. This leads to the computational assignment of 'hopped' reads to the incorrect sample, when combinatorial library indexes are used. Index hopping affects a small but significant fraction of reads (0.06% to 0.51% for MiSeq platforms) and has implications for downstream analysis (Farouni et al., 2020, van der Valk et al., 2020). The number of per-sample index-hoped reads is proportional to the total

number of reads contributed by that sample to the total sequencing pool. Often, a primary goal of BCR repertoire analysis is the detection of public (found in multiple individuals) antigen-specific antibody sequences, indicative of convergent response. Therefore, sequence sharing due to index hopping is an important concern (Yaari and Kleinstein, 2015, Galson et al., 2020, de la Rosa et al., 2017).

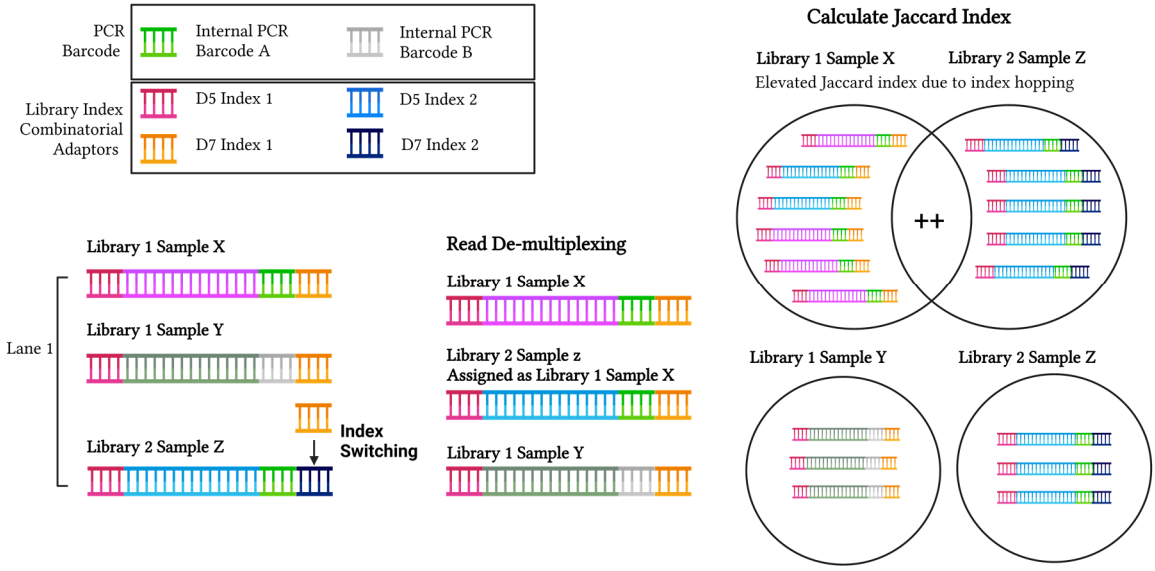


Figure 24. Schematic depicting computational detection of index hopping in sequencing data, due to the use of combinatorial library adaptors in conjunction with internal PCR barcodes.

The inclusion of an additional layer of barcoding during multiplex PCR allows for index hopping to be estimated. This is due to the enriched VDJ sequence sharing, measured by the Jaccard index, between patient samples that share an internal PCR barcode but belong to different libraries in comparison to samples that do not share an internal PCR barcode and come from different libraries.

Our BCR repertoire amplification protocol incorporates two layers of multiplexing. First, an internal sample-specific barcode (B1-B8) is incorporated into the DNA amplicon during PCR amplification [**Figure 12**]. Then, a single combinatorial library index is ligated to all sequences in a pool of up to eight multiplexed samples during library preparation. KAPA Dual-Indexed Adapters (KK8721) use eight P5 indexes and twelve P7 indexes in combination to generate 96 unique P5+P7 adaptors. When sequencing pools have high

plexity such as ours, this enables index hopping. Our double-indexing strategy means that within a sequencing lane, index hopping presents as enriched VDJ sharing between samples that share the same internal PCR barcode but different library barcodes [**Figure 24**]. Therefore, the unique VDJ sequence overlap for all pairwise combinations of samples was calculated using the Jaccard index [**Figure 25**].

The mean Jaccard index for related samples (defined as longitudinal samples from the same individual) was similar irrespective of whether samples were sequenced on the same or different sequencing lanes. The Jaccard index was decreased between related healthy volunteer samples compared to related sepsis samples, likely due to B cell clonal expansion in response to infection (Venet et al., 2013, Nielsen et al., 2020a) [**Figure 25**]. The marginal increase in Jaccard index between samples sharing an internal PCR barcode but separate library barcode when comparing those sequenced on the same versus separate lanes suggests low-level index hopping. The increase is minimal in comparison to the increased sharing between related sepsis samples, so it is unlikely to affect downstream analysis.

However, an increase in Jaccard index was seen between samples which shared internal barcodes and those that did not irrespective of whether they were sequenced on the same lane. This is indicative of either library pool or library adaptor contamination albeit at a low rate orders of magnitude lower than the sharing in related samples. Best practises were followed during experimental workflows and no contamination was visible on agarose gels for either RT or PCR negative controls. The low rate of contamination is unlikely to skew individual repertoire summary statistics but caution is required when investigating public BCRs. The use of a negative control library (H_2O) during library preparation would be beneficial in future studies, particularly given this is a complicated process involving multiple wash steps.

For future work, I recommend the use of KAPA Unique Dual-Indexed Adapters (KK8726), despite the substantial increase in cost. These consist of 96 unique adapters and indexes are never repeated throughout the set (*e.g.* 96 unique P5 indexes and 96 unique P7 indexes). This is more than sufficient for one lane of repertoire sequencing where no more than 80-90 samples from 10-11 libraries are multiplexed (Costello et al., 2018). With unique adaptors, index hopping produces 'phantom molecules' that cannot be assigned to any library and are subsequently excluded, reducing the risk of false-positive public clone detection (Farouni et al., 2020). They were not available to purchase at the time of the experimental work conducted in this thesis, but are now recommended by Illumina in place of dual combinatorial adaptors (Costello et al., 2018, illumina, 2023).

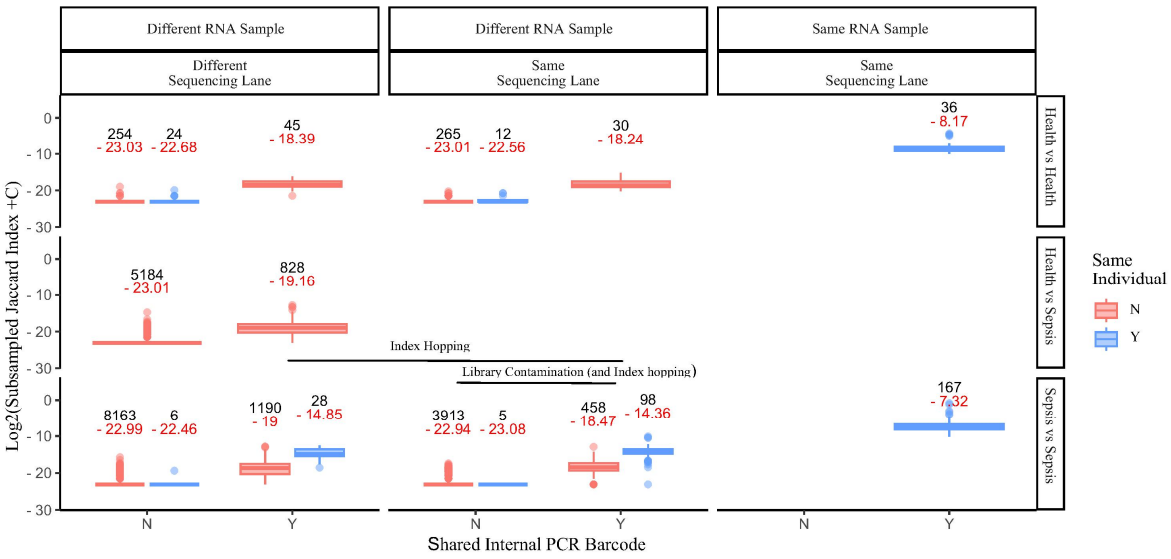


Figure 25. QC summary: unique BCR VDJ sequence overlap between pairs of samples.

The Jaccard index was used to investigate BCR repertoire sharing between all possible pairs of samples by subsampling each repertoire, calculating the sequence overlap using the Jaccard index and generating a mean of 10,000 repeats. For visualisation purposes the log2(Jaccard index+C) was calculated. C is a small constant equal to the minimum Jaccard index which does not equal 0, divided by 2. The addition of C was used to avoid log-0 error. Sample pairs are grouped according to whether samples were run on the same sequencing lane, had the same internal PCR barcode and whether they came from the same individual. As a reference scale, the Jaccard index was also calculated on two independent down-samples of the same repertoire. Even within the same sample, repertoire diversity and failure to reach sequencing saturation results in minimal VDJ overlap. For each boxplot, the number of pairwise comparisons are annotated in black; large differences in group size prevented statistical analysis. The mean for each group is reported to 2 decimal places in red. The median, first and third quartile are shown, while the whiskers represent 1.5 times the interquartile range (IQR).

3.4.4 Rarefaction Analysis

Rarefaction analysis was used to determine whether BCR sequencing depth was sufficiently large to estimate repertoire diversity without normalisation to account for differences in read depth. The mean number of unique BCR sequences was calculated for 20 subsamples of varying sizes, ranging from one to the maximum number of BCRs observed in each sample. The number of unique BCRs was plotted against the sample size [**Figure 26.A**]. In rarefaction plots, a curve means that sequencing depth has reached saturation whereas an ascending graph indicates insufficient read depth (Shugay et al., 2015, Chakraborty et al., 2022, Nakanishi et al., 2016). For all samples, sequencing saturation was not reached, hence measures of diversity are influenced by read depth. Consequently, best practise is to perform metric normalisation. Individual repertoires are down-sampled to the number of reads in the smallest sample and the diversity metric is calculated. This process is repeated to obtain the mean of resamples which accounts for differences in library size and stochastic sampling (Shugay et al., 2015). Many of the summary measures generated by the in-house BCR computational pipeline default to this normalisation method. Rarefaction analysis also revealed the significant clonal expansion in individual UK59580028, who was excluded from downstream analysis due to leukaemia (**3.4.1.1**) [**Figure 26.B**].

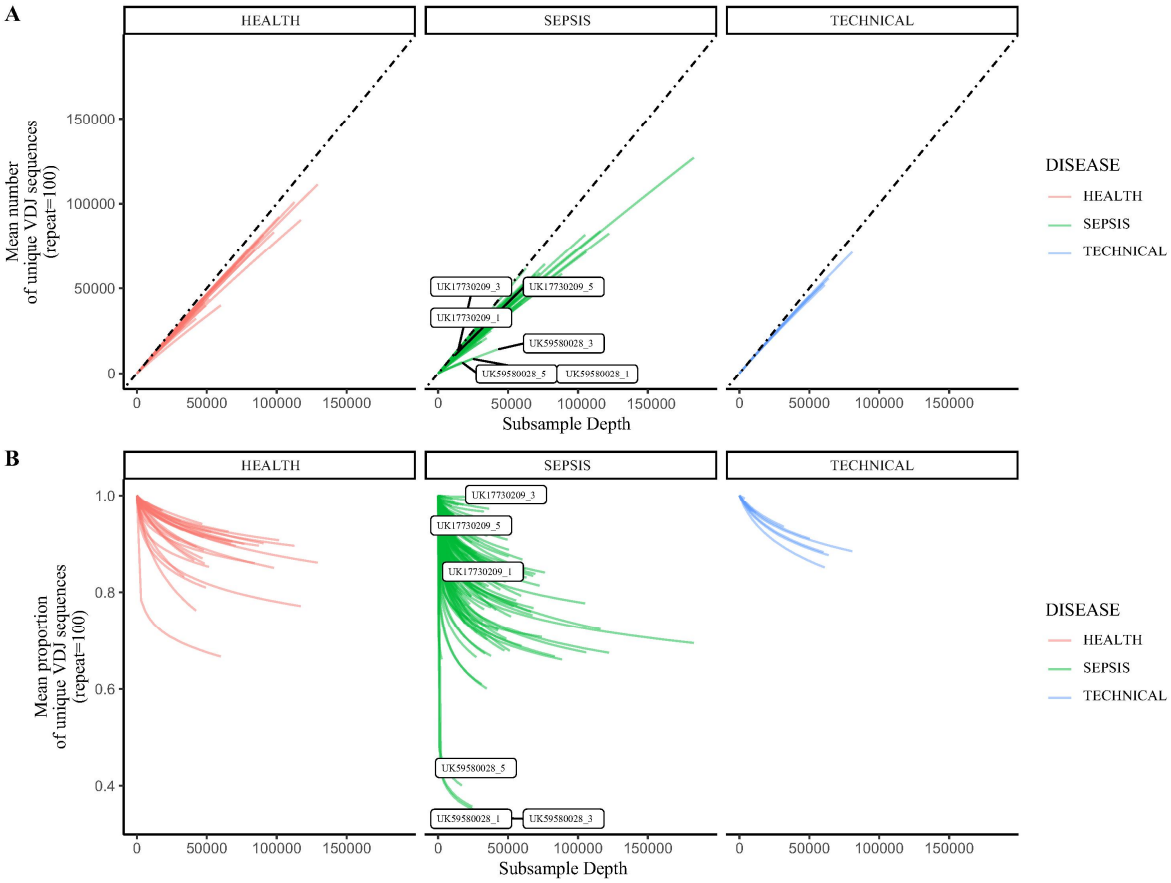


Figure 26. QC summary: BCR repertoire rarefaction analysis.

A) The mean number of unique BCRs in a subsample of the repertoire, plotted against the size of the subsample for 100 repeats per sample and 20 subsamples of increasing size. The black dotted line represents a one-to-one relationship. **B)** The mean proportion of unique BCRs in a subsample of the repertoire for 100 repeats is plotted against subsample size per sample. Sepsis samples generally exhibited a lower number of unique reads per subsample compared to healthy controls and the technical replicate. Individuals UK17730209 and UK59580028 are marked as they were excluded from downstream analysis due to significant comorbidities - metastatic tumour and leukaemia respectively. For the latter, substantial clonal expansion is clearly visible.

3.4.5 Quality Control Summary

Comparing quality control metrics across technical replicates suggests there was no substantial batch effect between sequencing lanes and libraries despite the significant period of time between amplification and sequencing of the first and last lanes. Downstream statistical analysis was therefore performed without batch correction.

QC analyses demonstrate the robustness of the in-house multiplex PCR amplification and sequencing protocol, and provide valuable guidance for future studies. High-quality VDJ

sequences were then annotated using IMGT HighV-QUEST. IMGT summary files were input into the second half of the pre-processing pipeline where a range of repertoire summary measures were calculated (**2.4.2.9**) [**Figure 27**].

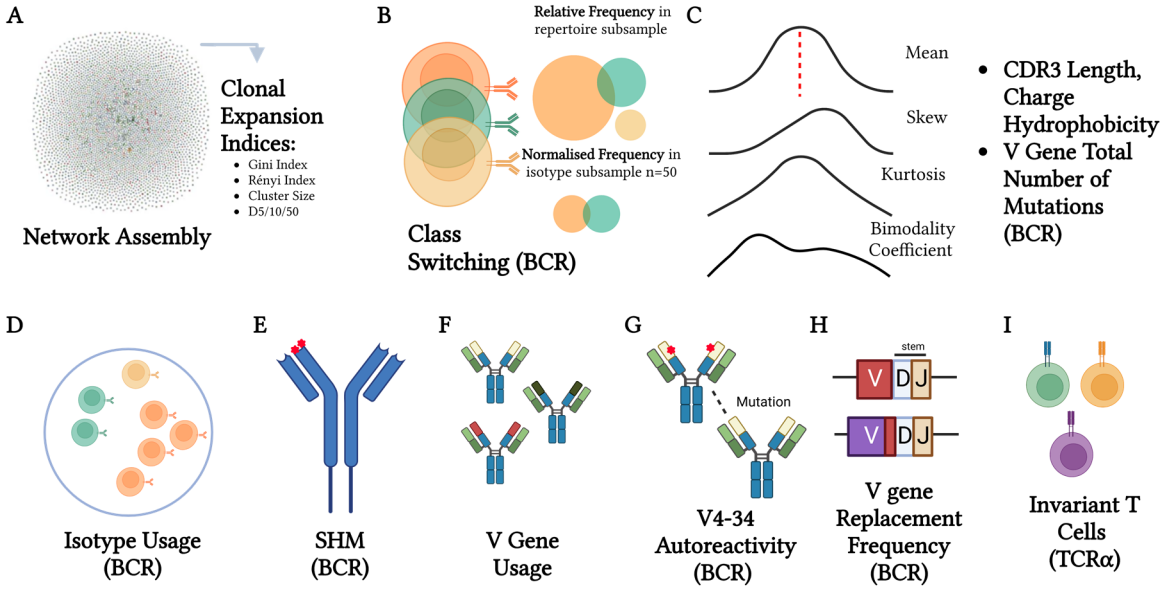


Figure 27. Repertoire summary metrics output by the in-house pre-processing pipeline.

A summary schematic detailing the range of different BCR and TCR repertoire summary measures output by the custom in-house processing pipeline. **A)** Clonal expansion indices calculated using the results of the network assembly stage. **B)** Two measures of class-switch recombination frequency. **C)** Statistical measures (mean, skew, kurtosis and the bimodality coefficient) which are used to describe the distribution of continuous traits such as CDR3 length. **D)** Total isotype usage. **E)** The rate of SHM. **F)** Isotype-specific V gene usage. **G)** B cell autoreactivity. **H)** V gene replacement. **I)** For the TCRA repertoire only: the proportion of invariant T cells.

3.5 Method Development: Dimensional Reduction of Repertoire Feature Matrices

3.5.1 Introduction and Aims

Dimensional reduction approaches, such as principal component analysis (PCA), have been widely used across different multi-omic data types for the reduction of large complex data matrices into a smaller number of components, which capture key axes of variation (Cantini

et al., 2021). By capturing both within and between-variable variation, these components can be used to ascertain features which distinguish between sample groups (such as disease versus health), and also identify novel patient subgroups in an unsupervised manner. Components can also be used to identify data-related issues such as batch effect. A comprehensive review of different methods is provided in Meng et al. (2016).

Within the field of adaptive immune repertoire analysis, dimensional reduction approaches are used to distinguish sources of biological variation between individual repertoires. For example Sevy et al. (2019) developed a method termed 'repertoire fingerprinting' which involves performing PCA on a normalised IGHV-IGHJ gene-pair count matrix. This technique effectively discriminated between the repertoires of HIV-infected plasmablasts after influenza vaccination compared to non-infected cells. Similarity, Kaplinsky et al. (2014) applied PCA to IGHV-IGHD-IGHJ rearrangement counts for IgH sequences from mice pre-B, follicular and marginal zone sorted B cell subsets. Their results show clear separation of cell types, driven predominantly by V_H rather than D or J_H gene usage. However, characterising the BCR repertoire solely on V-J gene usage encompasses only one aspect of the repertoire's complexity.

A more comprehensive repertoire dimensional reduction approach was implemented by Cortina-Ceballos et al. (2015). Multilevel PCA was performed on several IgH repertoire summary variables including diversity, IGHV gene usage and the rate of SHM for IgH sequences obtained from blood samples for volunteers immunised with either a trivalent or monovalent inactivated influenza vaccine. Visualisation of PC1 and PC2 successfully separated individual repertoires into two major clusters corresponding to vaccine type. Similarly, Kovaltsuk et al. (2020) and Galson et al. (2015a) used three different repertoire features: clonality, CDR3 length and V gene mutation rate for PCA dimensional reduction.

Again, the limited subset of repertoire features does not fully represent the complexity of the repertoire. However, their successes provide a promising foundation for more comprehensive analyses, such as demonstrated by Ghraichy et al. (2021). In this study, PCA was applied to 14 global repertoire metrics including V family and J gene usage, CDR3 junction length, diversity and the mean values for 9 CDR3 physiochemical properties including charge, hydrophobicity (GRAVY index) and polarity. Furthermore, global repertoire metrics were clustered using the k-means clustering algorithm to identify patient subgroups with similar repertoires. Importantly, care must be taken to ensure that input variables are independent of sequencing depth to avoid separation of patients based on technical factors (Chaudhary and Wesemann, 2018).

The output of our pre-processing pipeline is a BCR repertoire feature matrix comprised of more than 500 different BCR summary metrics calculated for both the total global repertoire and on each immunoglobulin isotype separately [**Figure 27**]. Consequently, data mining to identify disease-specific pathophysiology can be complex and incur a significant multiple-testing burden. Thus, there is the requirement to build on existing approaches and develop a novel computational method optimised for dimensional reduction of immune repertoire feature matrices. A successful method would generate a small number of informative components that capture the main axes of biological variation. In the following sections, I investigate two distinct dimensional reduction approaches to address this aim, using the sepsis BCR repertoire data to demonstrate their application. These are:

1. A custom pipeline based on the principles of weighted gene correlation network analysis (WGCNA) as published by Langfelder and Horvath (2008) and discussed in section **3.5.2**.

2. Tensor decomposition using Sparse Decomposition of Arrays (SDA) published by Hore et al. (2016) and addressed in section **3.5.3**.

3.5.2 Repertoire Weighted Gene Correlation Network Analysis (rWGCNA)

3.5.2.1 Introduction and Method Summary

WGCNA, developed by Langfelder and Horvath (2008), is a dimensional reduction approach specifically developed to reduce large gene expression count matrices, typically consisting of around 20,000 genes (in humans) into a smaller number of 'modules' which capture correlated gene expression networks. The process involves several computational stages outlined in detail in **Table 20**. Module scores are then explored in relation to an external trait of interest such as survival time or a binary trait like disease status, using two different approaches. One approach involves performing standard correlation or regression analysis between the module score and trait of interest. Genes with high module membership in modules that exhibit significant associations with the trait are interesting candidates for further investigation. An alternative approach is to calculate the significance of each gene in relation to the external trait which identifies biologically interesting gene candidates warranting experimental validation. Module significance is then determined by averaging the gene significance across the module, though this approach incurs a substantial multiple testing burden.

Stage	Description
1. Pre-processing	Samples and genes with high missing data are removed or, in some cases, missing data is imputed as PCA requires a complete matrix with no missing data.
2. Co-expression Network	The generation of a feature (gene) co-expression network using the Pearson correlation coefficient. Correlations between genes are the result of the interaction patterns between genes within and across biological pathways.
3. Module Identification	The identification of modules of highly correlated genes (features) using hierarchical clustering and Dynamic Hybrid Tree Cut.
4. Principal Component Analysis (PCA)	PCA analysis is performed on each cluster separately; the first principal component (PC1) for each cluster captures the maximal cluster variance and is termed the eigenvector (or eigengene). The eigenvectors for each cluster are combined into a matrix and are used as new sample coordinates for downstream statistical analysis.

Table 20. Weighted gene correlation network analysis (WGCNA) summary.

Table detailing the key computational stages in WGCNA analysis, as described in Langfelder and Horvath (2008). Each stage was subsequently optimised for repertoire WGCNA analysis.

The fundamental advantage of WGCNA compared to direct application of PCA, is the generation of a co-expression network [**Table 20**. **Stage 2**], which is then clustered. PCA is performed independently on each cluster and the first principle component (eigenvector) is used as the new module score [**Table 20**. **Stage 4**]. By first grouping correlated features, the co-expression network informs on shared regulation. As WGCNA is most commonly applied to gene expression datasets (RNA-seq or microarray), the correlated genes from significant modules can be input into gene ontology and pathway enrichment analysis software packages to identify the biological processes represented by the module and implicated in the trait of interest (Langfelder and Horvath, 2008, Peng et al., 2021).

WGCNA has, to date (2023), not been published for use in adaptive immune repertoire analysis. However, it has been used to identify gene modules which correlate to immune-repertoire diversity to inform on repertoire regulatory processes (Ye et al., 2022). Therefore, the work presented in this chapter has wider relevance to the adaptive immune receptor repertoire (AIRR) community.

The BCR repertoire feature matrix differs substantially from a gene expression count matrix, potentially impacting the performance of WGCNA. Firstly, it is orders of magnitude smaller than a typical gene expression matrix. Secondly, instead of uniform read counts the matrix is comprised of a range of different repertoire summary metrics (proportions, ratios *etc.*). Finally, there is missing data due to the low capture rate of rare antibody immunoglobulin isotypes and implementation of subsampling approaches. Considering these differences, I refactored and evaluated the performance of a custom pipeline that adapts the principles of WGCNA and is optimised for analysis of AIRR-seq data [**Figure 28**].

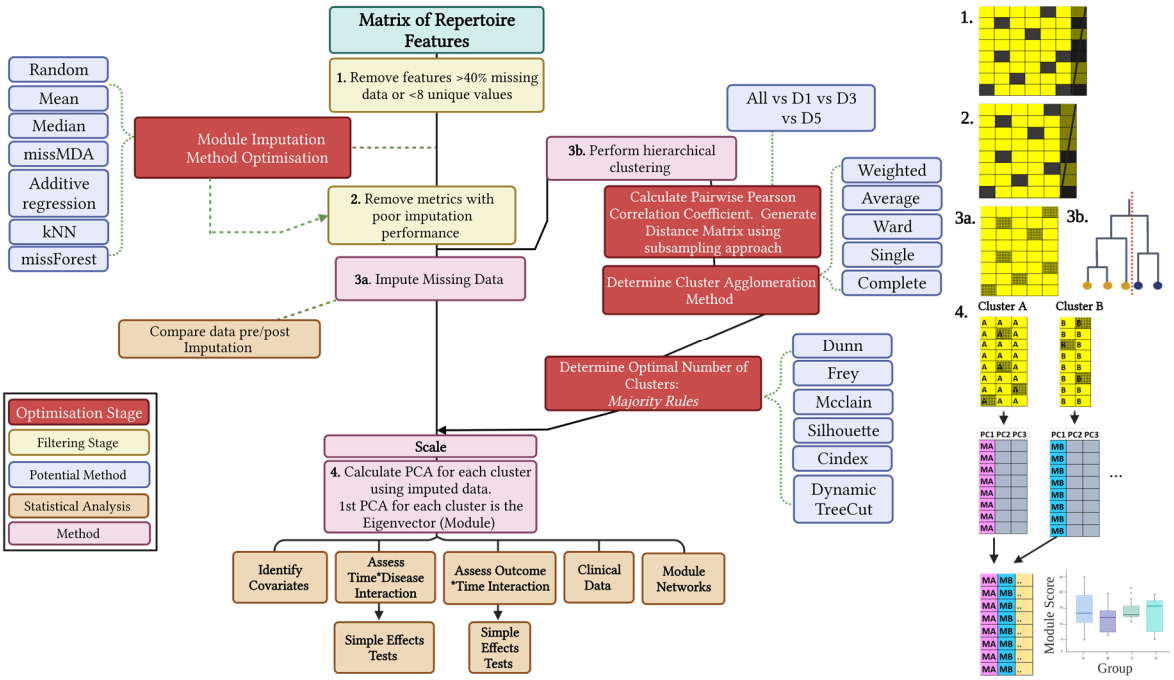


Figure 28. Custom rWGCNA dimensional reduction pipeline utilising the principles of weighted gene correlation network analysis (WGCNA).

Method stages are coloured according to theme. Optimisation stages are shown in red, with the corresponding exploratory methods highlighted in blue. These were evaluated in the repertoire dimensional reduction pipeline to determine the optimal method. Downstream statistical analyses are shown in orange and include identification of modules significantly associated with sepsis and sepsis mortality. To the right of the flow chart the different stages 1-4 are visualised in terms of the final repertoire feature matrix with missing data highlighted in black, and the output matrix of module scores which is used for conventional statistical analysis.

3.5.2.2 Imputation of Missing Immune Repertoire Data

3.5.2.2.1 Method Optimisation

PCA requires a complete numerical matrix with no missing data. Missing data occurred in the BCR repertoire feature matrix due to the down-sampling approach used to calculate per isotype summary statistics and account for variations in read depth (**3.4.4**). Where the productive read depth for a specific isotype did not exceeded the user-specified subsampling depth within an individual, a missing value was generated and represented as '*NA*'.

In the default WGCNA pipeline, samples or features with high missing data are either removed or replaced with a non-missing constant during pre-processing (Langfelder and Horvath, 2007, Langfelder and Horvath, 2008) [**Table 20.Stage 1**]. Similarly, the BCR repertoire matrix was first filtered, removing features with missing data exceeding 40% and/or minimal variation (<8 unique values). After filtering, the BCR matrix contained 101 features with missing data greater than 0% and less than or equal to 40%. In order to retain these features, the remaining missing data was imputed.

As the performance of different imputation methods for use with repertoire data has not been previously assessed, eight common imputation methods were evaluated. These were: mean, median and random imputation, and the more complex approaches, additive regression, regularised and expectation-maximisation PCA imputation, k-nearest-neighbours imputation (kNN), regularised imputation and random forest (RF) imputation. Of these, only KNN is available as part of the WGCNA R package. A summary of the methods can be found in **Table B. 4**.

To determine the optimal imputation method, a 'complete-cases' repertoire feature matrix was generated using the subset of samples that had no missing data. The completed matrix

was then populated with '*NA*' values at set frequencies (ranging from 0.05 to 0.5 at 0.05 intervals) using the *prodNA()* function from the R package 'missForest'. Initially, missing data was seeded feature-wise, and imputation, method selection and feature filtering was performed. This was then repeated with '*NA*s' seeded sample-wise to conduct feature filtering. Seeding missing values column-wise or row-wise rather than across the entire matrix ensured that every sample or repertoire feature had the same frequency of missing data.

To assess the performance of each imputation method, the root mean squared error (RMSE) for each feature was calculated by comparing the original values with the imputed values [**Equation B. 1**]. Normalised RMSE scores for feature-wise NA seeding are reported in **Table B. 5** [**Equation B. 1**]. Across all '*NA'* seed frequencies 'missForest' (RF) consistently achieved the lowest mean RMSE, performed significantly better than the base-mean and was chosen for downstream analysis [**Figure 29**]. At low '*NA*' seed frequencies (≤0.1) kNN imputation performed similarly to missForest. However, at greater '*NA*' seed frequencies (0.1-0.5) it was out-performed by RF (missForest) imputation. This can be attributed to the robustness of missForest in handling noisy data, including weak and non-informative predictors, which pose challenges for methods like kNN (Stekhoven and Buhlmann, 2012).

3.5.2.2.2 Feature Filtering

To ensure imputation was reliable, the feature-specific RMSEs obtained from imputation optimisation (**3.5.2.2.1**) using missForest were used to identify features with poor imputation performance and exclude them. An RMSE of ≤0.5 was considered acceptable. The proportion of missing data for each feature in the original repertoire feature matrix was

compared to the RMSE values calculated for that feature in the column-wise *NA* seeded datasets at various *NA* seed frequencies. If the proportion of missing data for a specific feature in the BCR matrix was larger than the *NA* seed frequency at which the feature-specific RMSE exceeded 0.5, that particular feature was removed from the matrix. Similarly, if the RMSE for a feature exceeded 0.5 when the sample-wise *NA* frequency was greater than the proportion of missing data for any individual sample, the feature was discarded. Applying these strict criteria, two BCR features were removed from the matrix due to poor imputation: the percentage of the repertoire explained by the top five clones for both IgE and IgG4.

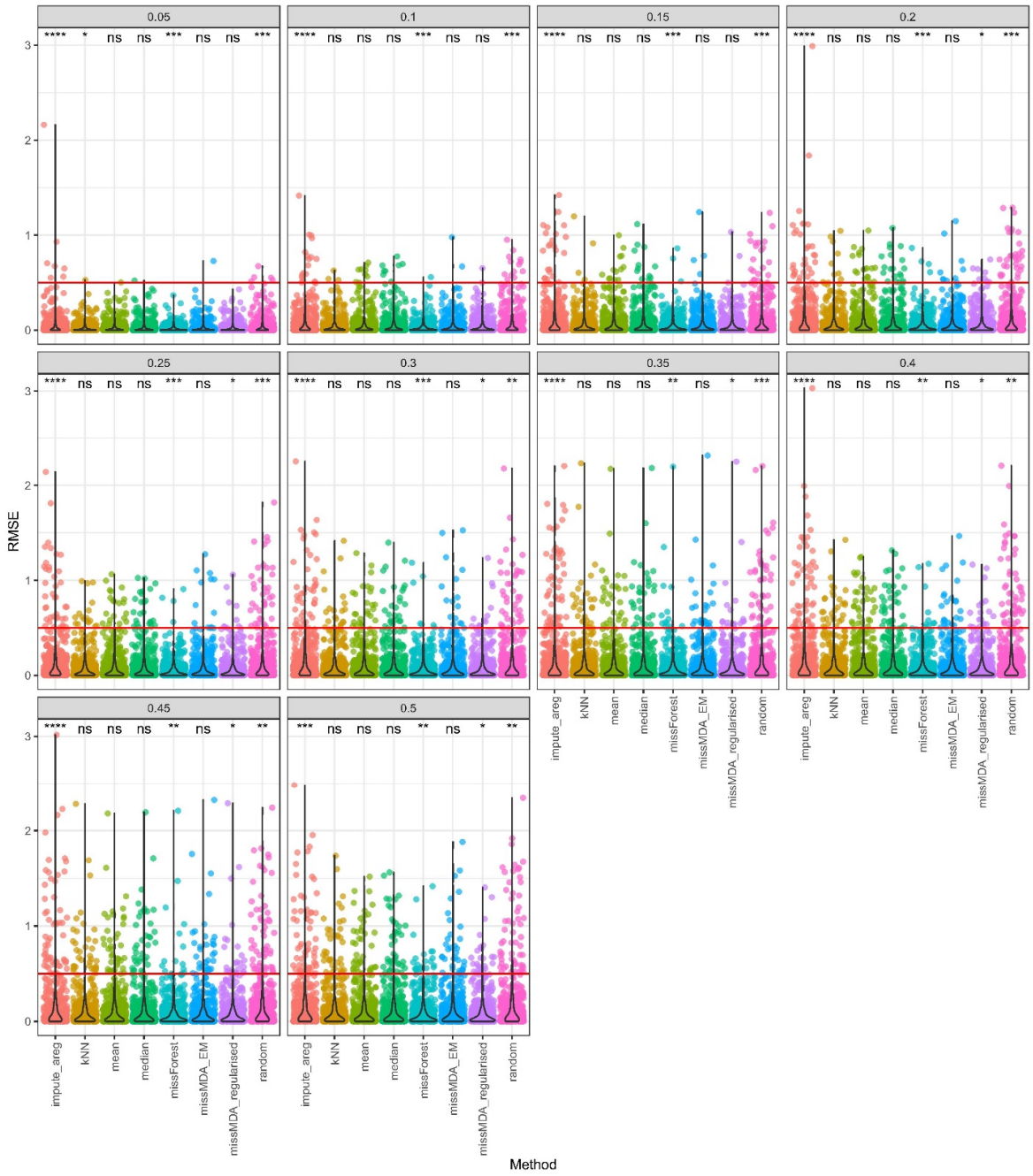


Figure 29. BCR repertoire missing data imputation optimisation.

Missing data imputation methods were evaluated by comparing the root mean squared error (RMSE) for all repertoire metrics at a given NA seed frequency for the BCR repertoire matrix. Results shown are for column-wise (feature) 'NA' seeding. Significance was calculated using a Wilcoxon Test comparing each method to the base-mean of all other groups for each NA seed frequency. The red line indicates RMSE=0.5 which is commonly accepted as 'good' performance. MissForest consistently outperformed the other imputation methods.

3.5.2.2.3 Repertoire Imputation Results

The missing data found in 101 features of the filtered BCR repertoire matrix was imputed using MissForest. A Wilcoxon test (assuming unequal variance) was used to assess whether there was a significant difference in feature mean after imputation. Out of the 101 features, only one (the percentage of the repertoire comprised of un-mutated IGHE) exhibited a significant difference in mean after imputation ($p<0.05$). For the majority of features, the absolute percentage change in mean after imputation was less than 10% [**Figure 30.A-C**]. There was a strong positive correlation between the number of missing values per feature and the absolute percentage change in mean after imputation ($R=0.57$, $p<0.05$), demonstrating the importance of pre-excluding features with high levels of missing data [**Figure 30.D**]. The imputed feature matrix was then scaled.

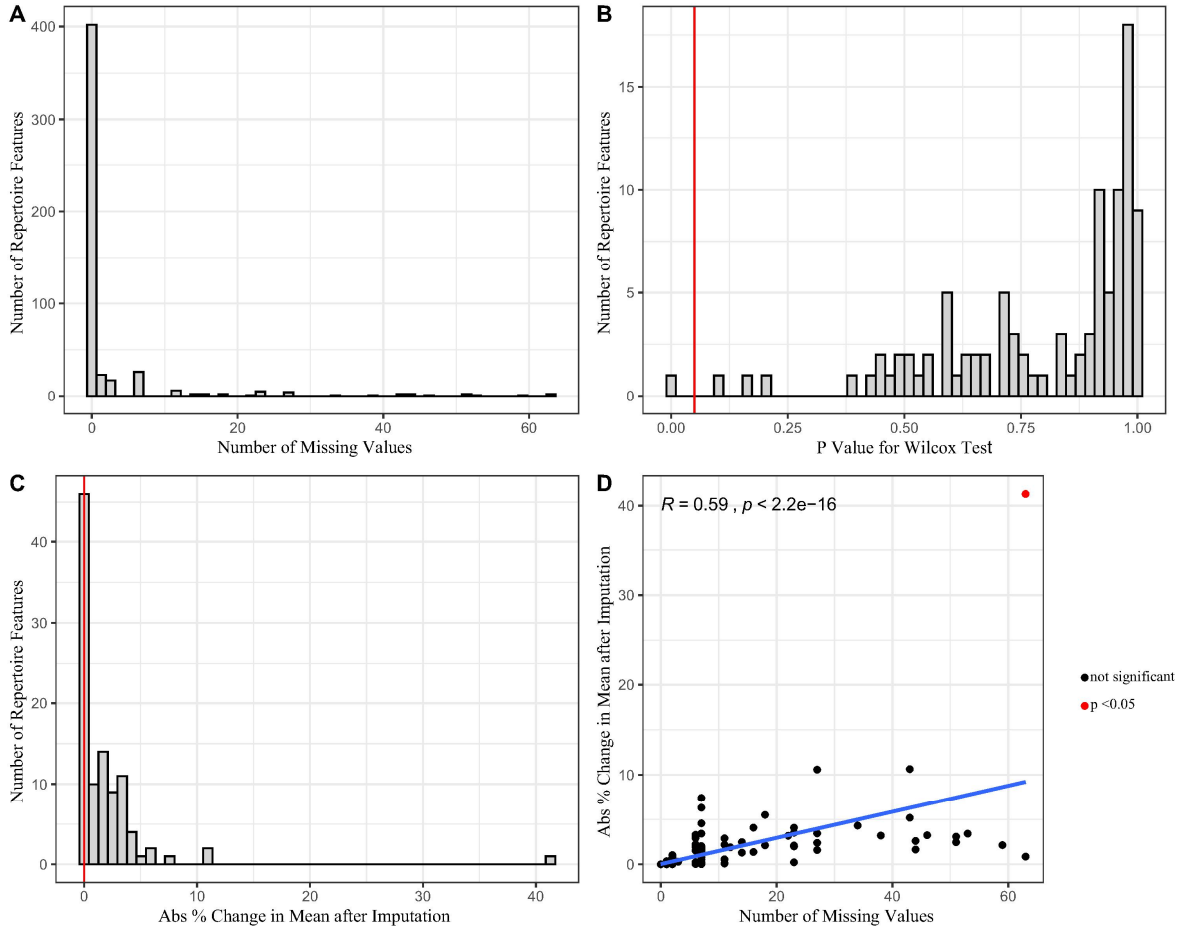


Figure 30. BCR repertoire imputation performance summary.

A) The number of missing values for each of 503 BCR repertoire features that passed basic filtering. 101 features had ≥1 missing values and required imputation. **B)** A Wilcoxon Test was used to identify a significant difference in mean after imputation with missForest. The mean percentage of unmutated IGHE was significantly different after imputation ($p<0.05$). **C)** The absolute percentage change in mean post-imputation for features with missing data. For 99% of features the absolute percentage change in mean was <10%. **D)** There was a significant positive correlation between the number of missing values and the absolute percentage change in mean ($R=0.57$, $p<0.05$).

3.5.2.3 Generation of the Distance Matrix

3.5.2.3.1 Accounting for Missing Data

The BCR repertoire feature correlation matrix was calculated using the Pearson correlation coefficient (PCC) applied pairwise to all features, prior to imputation of missing data to avoid the influence of any inaccurate imputed values. To account for missing data, each pairwise PCC was calculated on a down-sample that contained no missing values. This was

repeated 10,000 times, and the average Pearson correlation was calculated for each pair of features. The subsampled depth was set to a value of 0.8 multiplied by the minimum number of shared individuals across all pairwise comparisons (n=88). The correlation matrix was transformed into a distance matrix using **Equation B. 3**.

3.5.2.3.2 Investigating the Effect of Unbalanced Paired Design on Co-expression Networks

The study presented in this thesis employs a longitudinal paired design and it is important to consider the non-independence of samples during statistical analysis. When calculating the Pearson's correlation coefficient (PCC) using all available temporal data, as described in **3.5.2.3.1** and **Table 20**. **Stage 4**, the assumption of independence is violated. This can cause misleading results due to differing correlations within versus between participants (Bakdash and Marusich, 2017). There is no consensus in the literature regarding the application of Pearson's correlation for longitudinal data within the weighted gene correlation network analysis (WGCNA) pipeline. Several studies have used WGCNA on time course data irrespective of this statistical violation. Indeed, analysis of non-independent data using techniques that assume independence is a widespread practise (Wu et al., 2022b, Bakdash and Marusich, 2017).

Li et al. (2018) attempted to address this issue, and concluded that a PCC network derived from all samples, regardless of paired design, enables detection of biologically meaningful WGCNA modules. However, the authors emphasise the need to account for non-independence of samples in downstream analysis by using linear mixed effect (LME) models. However, they did not specifically address the implications of unbalanced experimental design (unequal number of observations per category) on co-expression networks.

In the case of the BCR repertoire cohort, longitudinal data is truncated by death, hence individuals have varying numbers of repeated observations. I hypothesised that using all samples to generate the co-expression network as suggested by Li et al. (2018) would bias network structure towards either within-participant correlations for sepsis survivors and health, as these individuals have the most samples, or day 1 dynamics as these samples constitute the largest proportion of the dataset. To explore this possibility, the correlation matrix was generated using either I) all available data or alternatively using samples from II) day 1 III) day 3 and IV) day 5 only. The subsample depths for each matrix were 88, 29, 28 and 27 respectively.

As expected, the pairwise PCCs derived using all data were most similar to those obtained from day 1 only ($R=0.92$, $p<2.2e^{-16}$) and less well correlated to those derived on day 5 only ($R=0.86$, $p<2.2e^{-16}$). This indicates that the co-expression network is a biased towards the group representing the largest number of observations per category [**Figure 31.A-C**]. When considering correlations derived from a single time point, adjacent time points were more correlated than those further apart (day 1 versus day 5). This suggests dynamic changes during the course of sepsis [**Figure 31.D-F**]. Alternatively, it reflects sample dropout due to death, with the day 5 network capturing co-expression patterns in sepsis survivors and healthy individuals.

Given the strong correlation between the PCCs derived from day 1 and those derived from all time points ($R =0.92$, $p<2.2e^{-16}$), I tested whether using only day 1 samples could function as a more statistically correct method to derive the correlation network in this longitudinal cohort. To assess this, I calculated ΔR, which is the change in PCC (R) for a pair of features when comparing two different methods. However, as shown in **Figure 32,** the distribution of ΔR when comparing day 1 PCCs versus either day 3 or five PCCs, had

a larger standard deviation compared to the comparison of PCCs derived from all time points against each individual time point - meaning that the day 1 co-expression network captured later time points more poorly than the one derived on all samples. This had implications for downstream clustering. In some cases, the trajectories of AIRR features which were assigned to the same cluster based on day 1 co-expression networks, diverged over the measured timeframe and the module eigenvector was unrepresentative. This effect was minimised when using PCCs derived from all time points and modules predominantly contained features with similar trajectories [**Figure B. 1**]. Therefore, the decision was made to use all available time points to generate the repertoire co-expression matrix, irrespective of the violation of statistical non-independence. These findings further support the work presented in Li et al. (2018) and offer valuable insights for researchers working with complex longitudinal data.

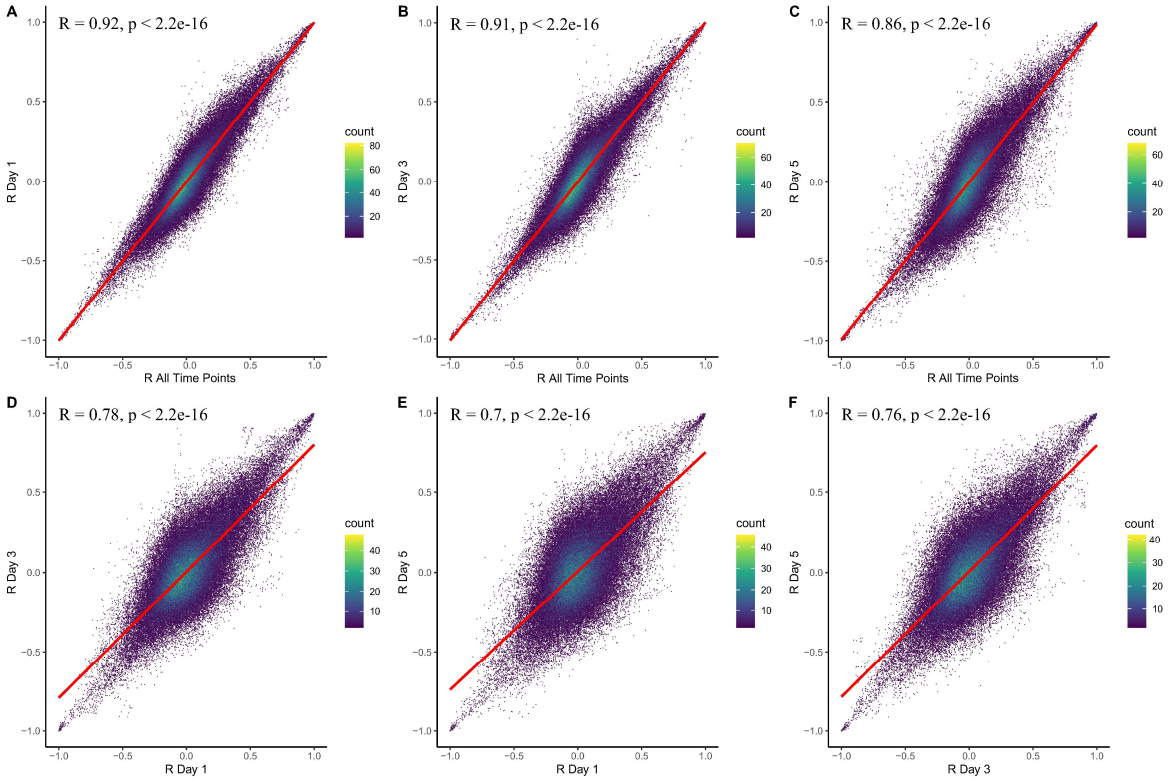


Figure 31. Comparing correlation matrices derived using all data versus the subset of data for each time point (day 1, 3 and 5) for the BCR repertoire matrix.

A-F) Each point represents a unique BCR AIRR feature pair. X and Y-axis are the PCC (R) for co-expression networks derived using either all time points or day 1, 3, and 5 only. A linear regression line is shown in red with corresponding R value; R ≠1 indicates a change in the relationship between BCR repertoire measures, either due to the effect of time or sample dropout due to death.

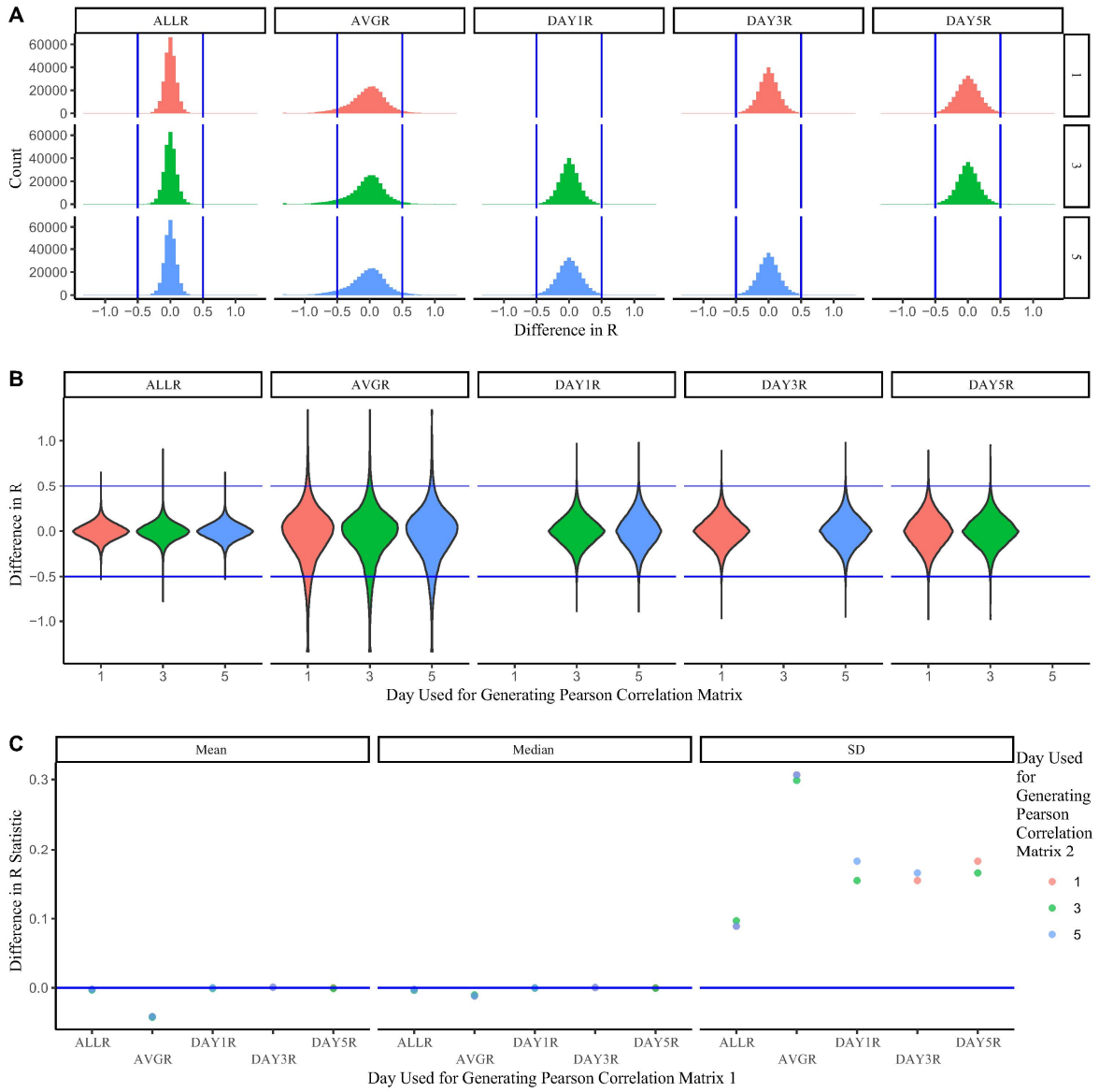


Figure 32. Comparing co-expression networks derived using PCCs for day 1, day 3, day 5, all available data (ALLR) and an average of day 1 and day 5 PCCs (AVGR).

The change in PCC (ΔR) for every pair of features was calculated for each pair of methods. **A)** Histograms showing the distribution of ΔR between pairs of methods. A good method would produce a ΔR distribution centred on 0 (no difference in correlation) with minimal spread. Where the same method is compared ΔR=0 for all feature pairs and has not been visualised. **B)** Violin plots conveying the same information as in A. **C)** The mean, median and SD for ΔR between different methods.

3.5.2.4 Hierarchical Clustering: Linkage Method

The next step in the WGCNA pipeline involves agglomerative (bottom-up) hierarchical clustering using the distance matrix from **3.5.2.3**, to identify groups of features that are co-

expressed. The choice of linkage method is important as it determines how the distance between sets of observations is calculated and can influence cluster structure.

I assessed the performance of five common linkage methods: weighted, average, ward, single and complete linkage using the agglomerative coefficient (AC) score which was calculated using the 'AGNES' R package. AC scores range from 0 to 1, values closer to 1 indicate a stronger clustering structure with well-defined clusters. Among the evaluated methods, Ward's method maximised the AC score (AC=0.97) and was used in agglomerative hierarchical clustering [**Figure 33**]. This result aligns with the default WGCNA settings. For a summary of the linkage methods, refer to Seyler et al. (2015) and Appendix **Table B. 6**.

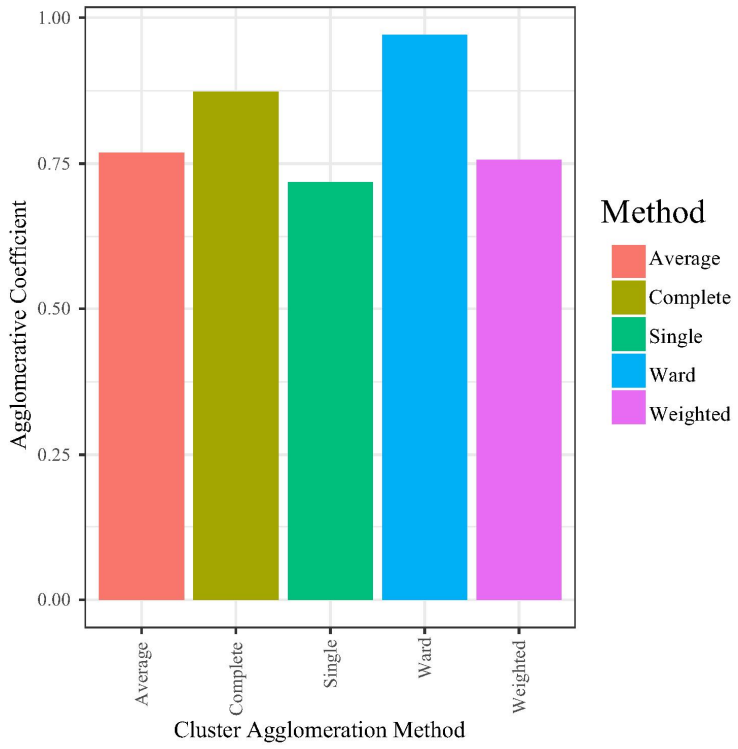


Figure 33. Determining the optimal linkage method for the BCR repertoire distance matrix using agglomerative coefficient (AC) score.

Five common linkage methods were compared using the agglomerative coefficient (AC) from the Agnes R package. Ward's linkage method resulted in the strongest clustering structure, as measured by a large AC score, while single linkage performed comparatively poorly.

3.5.2.5 Repertoire Feature Cluster Detection

3.5.2.5.1 Fixed Height Cutting

To determine the optimal number of immune repertoire feature clusters the 'NBclust' R package was used to run the five cluster detection methods available for hierarchical clustering (Dunn, Silhouette, Mcclain, Cindex and Frey). The details of these methods are summarised in Appendix **Table B. 1**. The optimal number of clusters (K) was determined based on the 'Majority Rules', where the value of K proposed by the highest number of methods is the best choice. Where multiple K values were tied, the largest value of K was chosen to achieve greater resolution.

I also examined the impact of the user-defined minimum and maximum cluster number parameters on the proposed optimal number of clusters. Two scenarios were observed [**Figure 34**]. Firstly, these cluster detection methods have a tendency to suggest the same number of clusters as the user-specified upper limit when using repertoire data. Secondly, when providing a low lower bound (<11), two was often suggested as the optimal number of clusters. This division aligns with the initial major split in the dendrogram and is likely driven by class-switched versus non-class-switched measures and inverse statistical measures [**Figure 35**]. Due to the instability in suggested K values proposed by the fixed height cutting methods, an automated pipeline implementing these methods is not feasible. Therefore, an alternative approach known as Dynamic Hybrid Tree Cut was investigated.

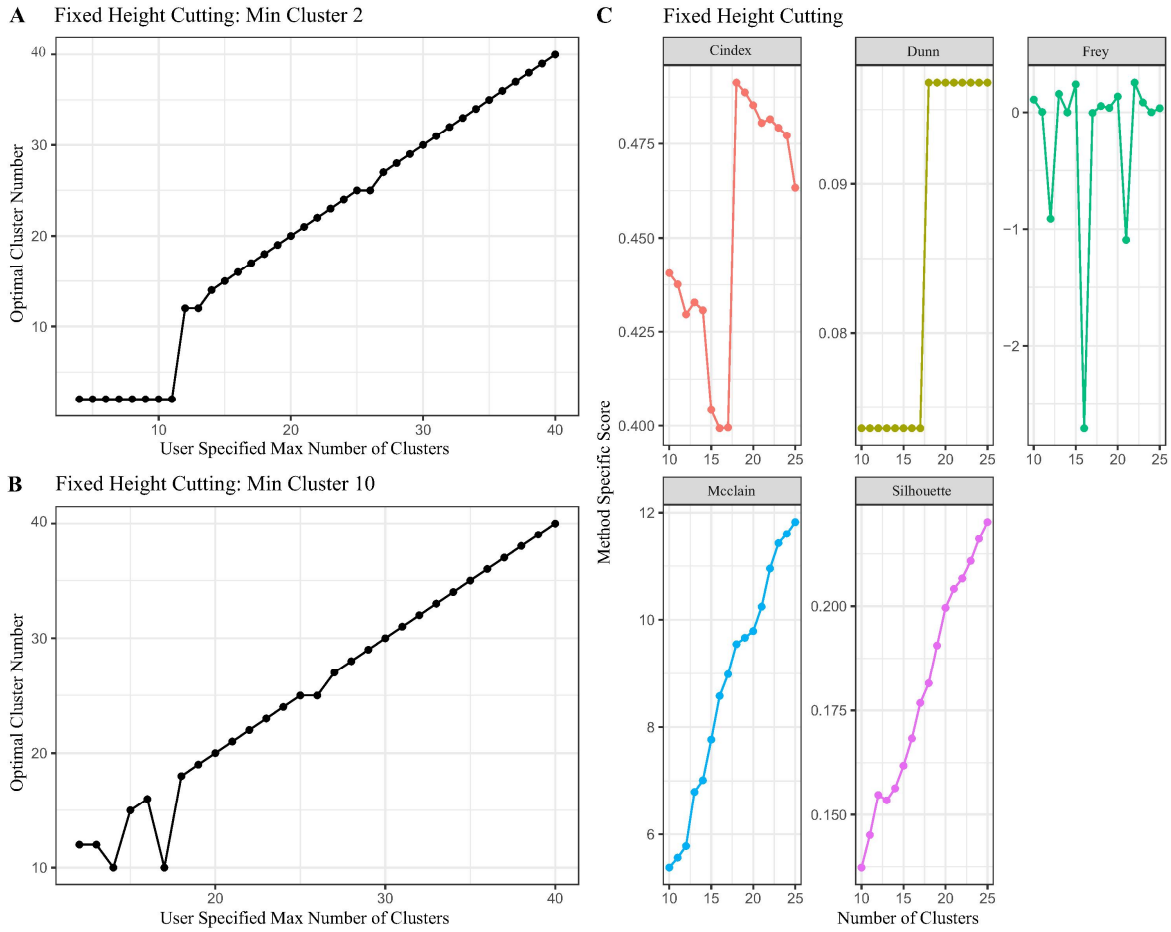


Figure 34. Determining the optimal number of clusters for the BCR repertoire distance matrix using fixed height cutting.

A,B) The optimal number of clusters proposed by the five cluster detection methods available in the 'NBclust' package for hierarchical clustering, for a variable maximum and a minimum number of clusters set to **A)** 2 and **B)** 10. **C)** Scores obtained from the five cluster detection methods at different values of K, where the parameters were set to a minimum number of clusters of 10 and a maximum of 25.

3.5.2.5.2 Dynamic Hybrid Tree Cut

Langfelder et al. (2008) proposed a new method for cluster detection termed Dynamic Hybrid Tree Cut (DHTC). The DHTC method utilises both the dissimilarity information between objects (distance matrix) and specified linkage method to build clusters bottom-up using branch shape to partition clusters. The branch shape is defined by four criteria: 1) a cluster must have a user-specified minimum number of member objects, 2) objects that are distant from the cluster are excluded, even if they fall on the same branch of the

dendrogram, 3) each cluster must be sufficiently separated from its surroundings by a gap and 4) the core of the cluster should be tightly connected (Langfelder and Horvath, 2008, Langfelder et al., 2008).

DHTC has been shown to out-perform fixed height cutting when dealing with data that has nested or complex structure, as in the case of the BCR repertoire data. Application of DHTC identified 31 BCR repertoire clusters (using a minimum number of features per cluster of 10) [**Figure 35**]. Six features were excluded as they were not sufficiently similar to the core features on their respective branches. Importantly, DHTC was able to recognise the nested structure arising from class-switched versus non-class switched isotype features without user-input, an improvement on fixed height cutting (**3.5.2.5.1**).

However, the proposed number of clusters remained sensitive to the user-specified parameter 'minimum features per cluster' and 'resolution'. When using a threshold of at least 11 features per cluster, a branch consisting of 10 IgG3 related measures was excluded. Thus, parameter choice was a compromise between maintaining features and generating a manageable number of biologically plausible clusters for downstream analysis.

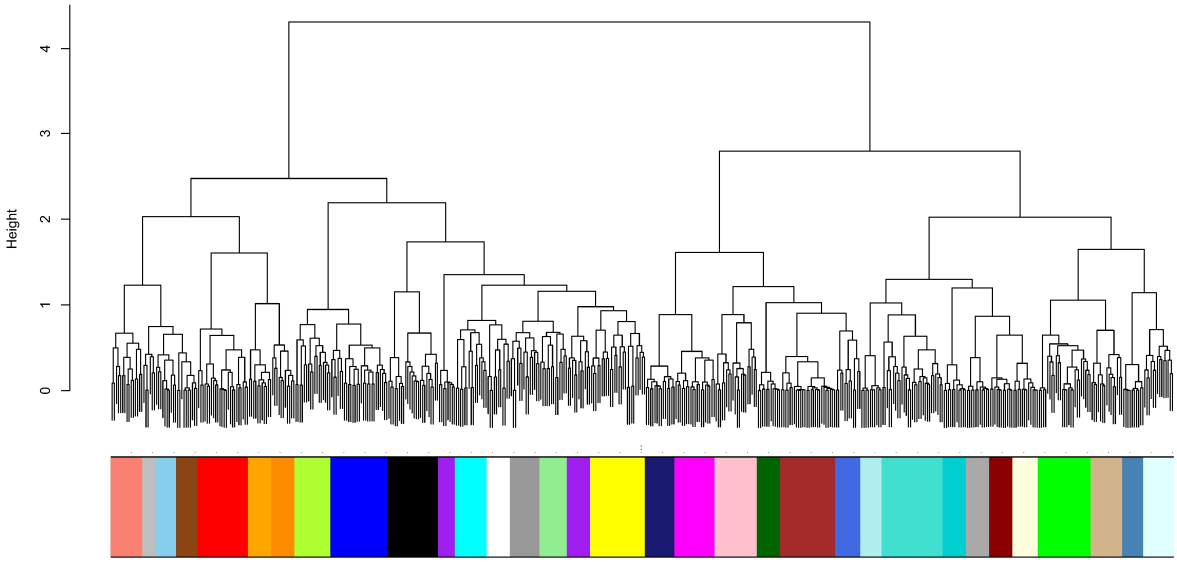


Figure 35. Dynamic Hybrid Tree Cutting.

The BCR repertoire feature dendrogram was generated using agglomerative hierarchical clustering specifying Ward's linkage, applied to the BCR repertoire distance matrix generated using all available samples. BCR feature labels have been omitted for visualisation. Module detection was performed using Dynamic Hybrid Tree Cut from WGCNA and identified 31 BCR repertoire clusters. Light grey indicates the 6 features which were excluded from a cluster due to dissimilarity to the other features on their branch.

3.5.2.6 Principal Component Analysis

3.5.2.6.1 Variance Explained by PC1

PCA was performed separately on each cluster identified in **3.5.2.5.2**, using the imputed, scaled BCR repertoire feature matrix and the '*prcomp()*' function [**Figure 36**]. The first principal component (PC1) is known as the eigenvector and was extracted as the new module score for each cluster, resulting in a matrix of x eigenvectors by n samples, where the number of eigenvectors (x, modules) is equal to the number of clusters (K=31).

In approximately 50% of clusters, PC1 captured over 50% of the cluster variance [**Figure 37**]. In some modules, *e.g.* module 4, PCA was unable to reduce the data into meaningful components, and all PCs captured only a small proportion of variance. Often these modules contained IGHV gene usage and statistical measures (*e.g.* kurtosis, skewness and bimodality

coefficient); perhaps these features represent an inherent limitation of the approach as their distribution and values are highly personalised.

Previous work by Sevy et al. (2019) has shown that direct application of PCA to a matrix of IGHV-J gene usage, produces components which capture a significant proportion of variance. However, further investigation revealed that these PCs are often driven by a subset of IGHV genes with a strong signal; many IGHV genes contribute very little to the overall component (Sevy et al., 2019). In rWGCNA, IGHV genes were grouped into different modules based on co-expression, prior to dimensional reduction. Therefore, it is not unexpected that some modules contain IGHV genes with a strong signal captured well by the module eigenvector, (as in Sevy et al. (2019)), while other modules capture less informative IGHV genes and features that show weak and variable signal.

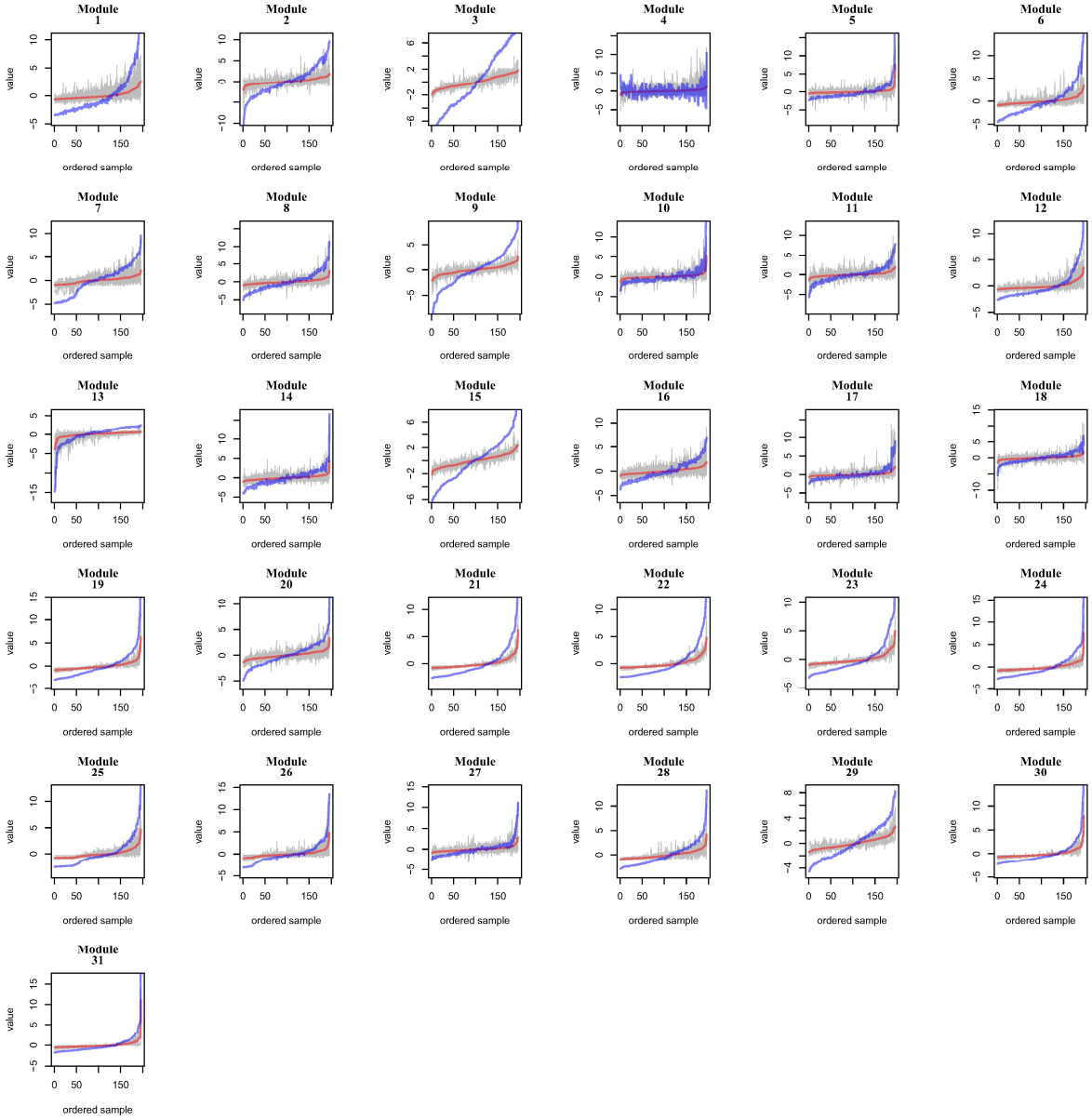


Figure 36. rWGCNA identified 31 BCR repertoire modules.

Plots show the scaled 'expression' for every feature within each of the 31 BCR modules, for every sample. The red line is the mean value across features for each sample. The blue line is the new module coordinate (PC1) for each sample. PCs are inherently directionless, but components were orientated so that they positively correlated with the mean for ease of interpretability, thus an increase in PC1 corresponds to an increase in mean 'expression'.

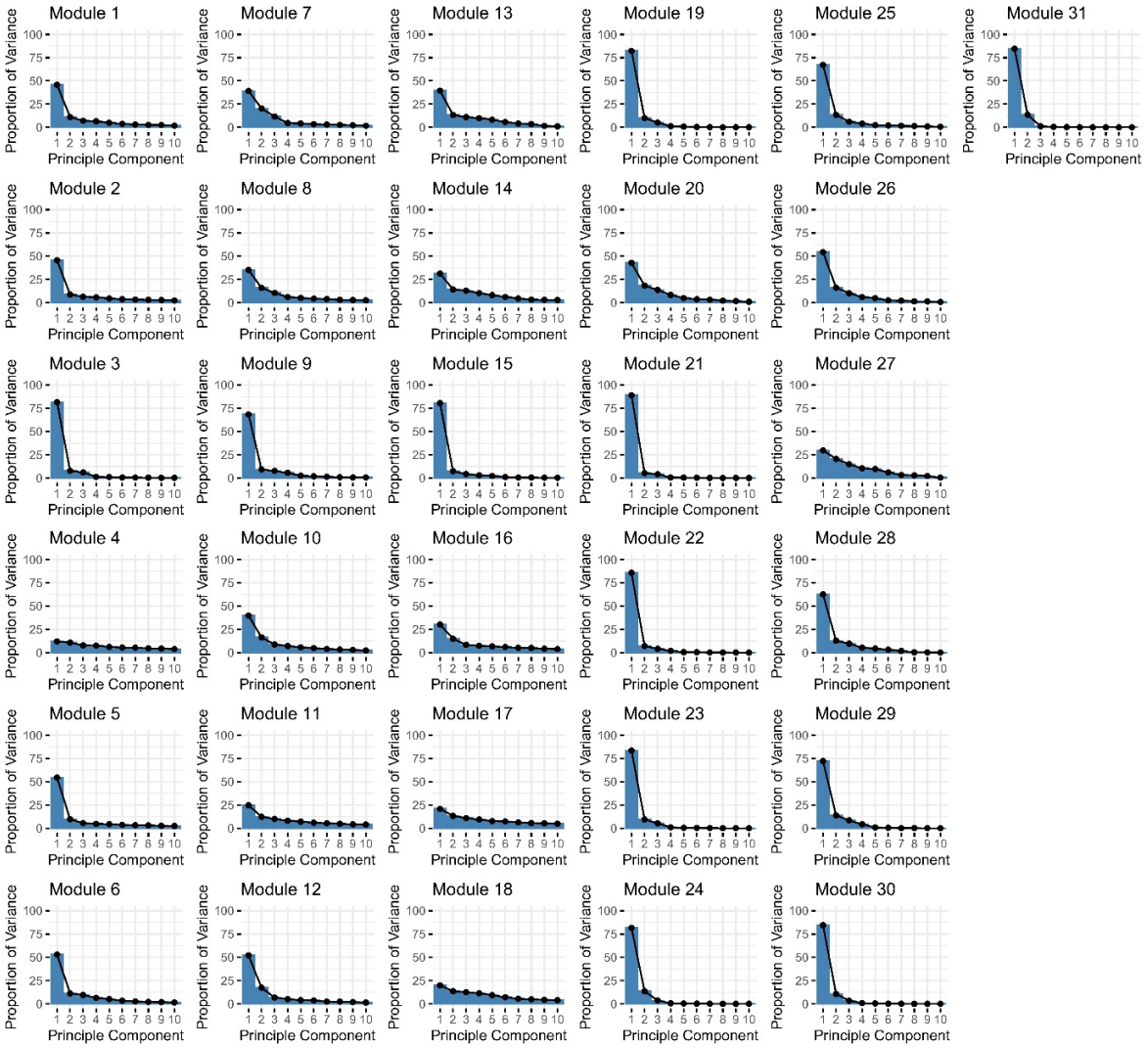


Figure 37. Proportion of cluster variance explained by each BCR eigenvector (module).

The proportion of variance explained by the first 10 principle components for each BCR module. 10 components are shown as each cluster had a minimum of 10 BCR repertoire features. For the majority of modules PC1 captures a large proportion of cluster variance and the new eigenvector coordinates well represent trends seen in individual features. For modules including 4, 11, 17, 18 and 27 PC1 captured very little variance. This may be due to the highly personalised use of IGHV genes.

3.5.2.6.2 Eigenvector Analysis: PC1 Loading Scores

For each cluster, the PC1 loading scores were also extracted and used to determine the extent to which each feature contributes to PC1 [**Figure 38**]. Large loadings indicate a strong association between the feature and component. Features which contribute more than expected to PC1, were identified by calculating the cluster specific loading score assuming all features contributed equally [**Equation B. 4**]. Absolute loading scores greater

than this value indicate a strong relationship and these features are good candidates for follow-up analysis in modules significantly associated with disease.

The distribution of loading scores within each module was compared, a large variance indicates that a subset of features drive the component, while other features are non-informative [**Figure 38**]. For the majority of BCR modules, features contributed equally to PC1 *e.g.* Module 3, hence all constituent features are well represented. This coincides with a high proportion of variance explained by PC1 [**Figure 37**]. Conversely, loading scores for modules where PC1 explained a low proportion of variance (4, 11, 17, 18 and 27) were variable, and a large proportion of features were under-represented. If these modules show a significant association with a trait of interest, caution is required when interpreting the underlying biology.

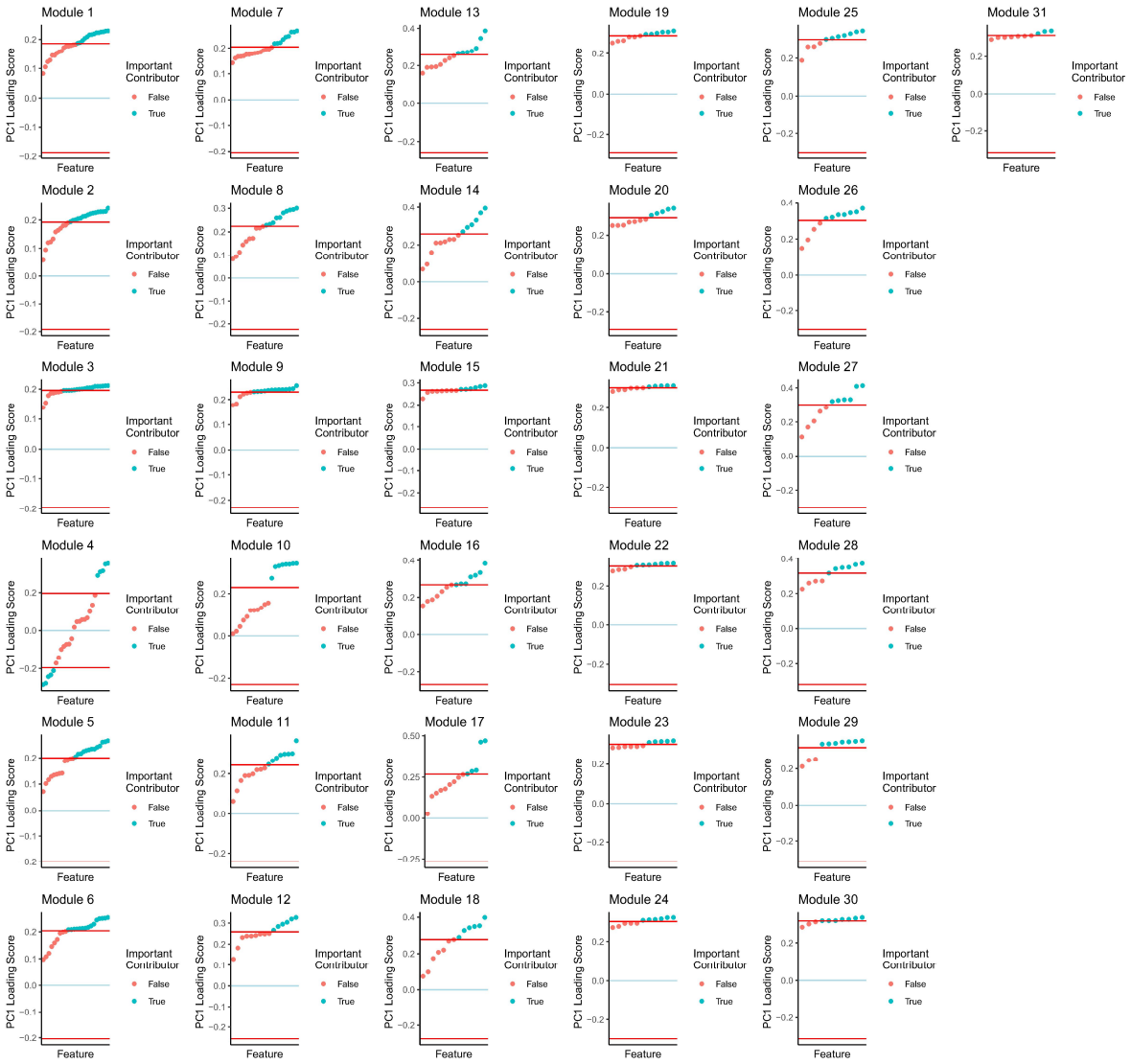


Figure 38. PC1 feature loading scores for the 31 BCR repertoire modules.

PCA loading scores are plotted for PC1 (the eigenvector) for each constituent feature, per module. The module specific 'large loading score' threshold is shown as the red horizontal line and features are classified according to this threshold. BCR repertoire features with large loading scores in modules associated with a trait of interest are interesting candidates for follow-up analysis.

3.5.2.7 rWGCNA Summary

In this section (**3.5.2**), I have described the development of an optimised pipeline for repertoire feature dimensional reduction by adapting the principles of WGCNA. Importantly, I show that missForest (Random Forest) can be used to reliably impute missing BCR repertoire data, and outperforms the WGCNA default of kNN. I also highlight

the effect of unbalanced paired design on correlation network analysis, which has not previously been addressed in the literature.

In total, 31 BCR IgH modules capturing correlated repertoire features were identified. Feature-loading scores can be used to identify interesting candidate features for further analysis. Module biology will be discussed in relation to sepsis pathology in **Chapter 4.**

3.5.3 Tensor Decomposition: Sparse Decomposition of Arrays

3.5.3.1 Introduction

To evaluate the robustness of the biological signals identified using rWGCNA, I applied a second dimensional reduction technique, Sparse Decomposition of Arrays (SDA) to the BCR feature matrix (Hore et al., 2016). Conceptually similar to rWGCNA, SDA uses Bayesian regression to reduce a tensor (a 2D or 3D data array) into a smaller number of latent components which capture major sources of variation. Components are then tested for statistical association with a trait of interest. In Hore et al. (2016) SDA was applied to an RNA-seq dataset comprised of matched adipose, lymphoblastoid cell line (LCLs) and skin gene expression data and successfully identified statistically significant gene networks with a genetic basis. SDA has not been published to date for use in adaptive immune repertoire analysis.

The latent components output by SDA consist of three scores, the sample and feature loadings and the Posterior Inclusion Probabilities (PIP). The sample loading score is a vector representing the magnitude of the effect of each component across individuals, analogous to the PC1/eigenvector (module score) in WGCNA. This is used as the new sample phenotype. The feature loading scores are conceptually equivalent to the PC1 loadings scores and measure the magnitude of effect for each feature within the component.

Finally, PIP scores are the likelihood that a feature loading score does not equal zero. SDA employs a sparse 'spike and slab prior' which enables the shrinkage of feature scores to zero. This makes it easier to identify the main features which are involved in the network captured by each component, as many are effectively removed from the model.

SDA offers several advantages over WGCNA. Firstly, there is no requirement for feature clustering. Instead, each SDA component consists of feature loadings for every feature found in the original matrix; features can contribute strongly and with high probability to the loading scores of multiple components allowing for complex interactions. Removing the cluster detection stage is beneficial because, as reported in **3.5.2.5**, it is highly sensitive to user-specified parameters. A second advantage is that while components are generally sparse (a small number of features with large loadings), the model allows for non-sparse components which arise due to confounding effects such as batch effect, where gene/feature expression is expected to be systematically shifted in a group of individuals. These components are identified by their strong correlation with sources of technical variation. Finally, the calculation of PIP scores provides an additional means to identify high confidence candidates from among all those included in a component, for follow-up analysis.

3.5.3.2 Identification of Robust Components

SDA was applied to the pre-filtered, imputed, BCR repertoire feature matrix from **3.5.2.2.3**. The number of components parameter was set to 100 as at values greater than this many all-zero components were returned. To identify robust components I followed the guidelines established in Hore et al. (2016) and ran ten iterations of SDA. Then, non-zero component sample scores from all runs (n=849) were correlated pairwise using the Pearson's Correlation Coefficient. Agglomerative hierarchical clustering was performed using Ward's

Linkage and it was terminated when no correlations between clusters were greater than 0.6 [**Figure 39**]. Within each cluster containing components from more than 5 runs of SDA (a robust signal), components were averaged and the mean feature, sample and PIP scores were calculated as new overall scores. This process identified 41 robust components detected in ≥5 iterations of SDA, for downstream analysis. Two types of components were identified, those containing unidirectional loading scores (*e.g.* component 7), which are simple to interpret and those with bidirectional loading scores (*e.g.* component 15) which are more challenging to interpret [**Figure 40**].

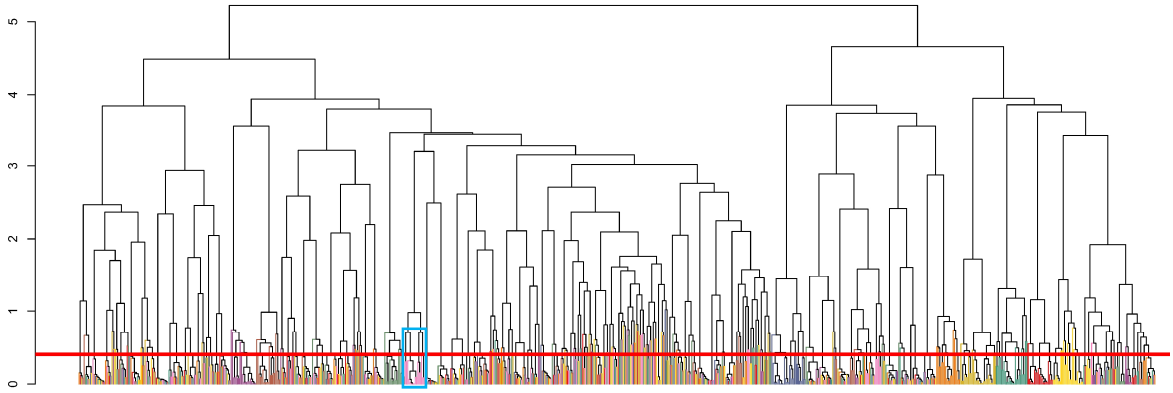


Figure 39. Agglomerative hierarchical clustering of the 849 non-zero components from 10 iterations of SDA, applied to the imputed BCR repertoire matrix.

Clustering was terminated when no correlations between clusters were ≥0.6 and components were grouped into clusters using fixed height cutting. Branches of the dendrogram are coloured according to SDA run (1-10). Some clusters are comprised of only components identified in one run (example outlined in blue); these are not robust biological signals. Forty-one clusters contained components from ≥5 iterations of SDA and the feature and sample loading scores and PIP scores from constituent runs were averaged within each cluster to generate robust components.

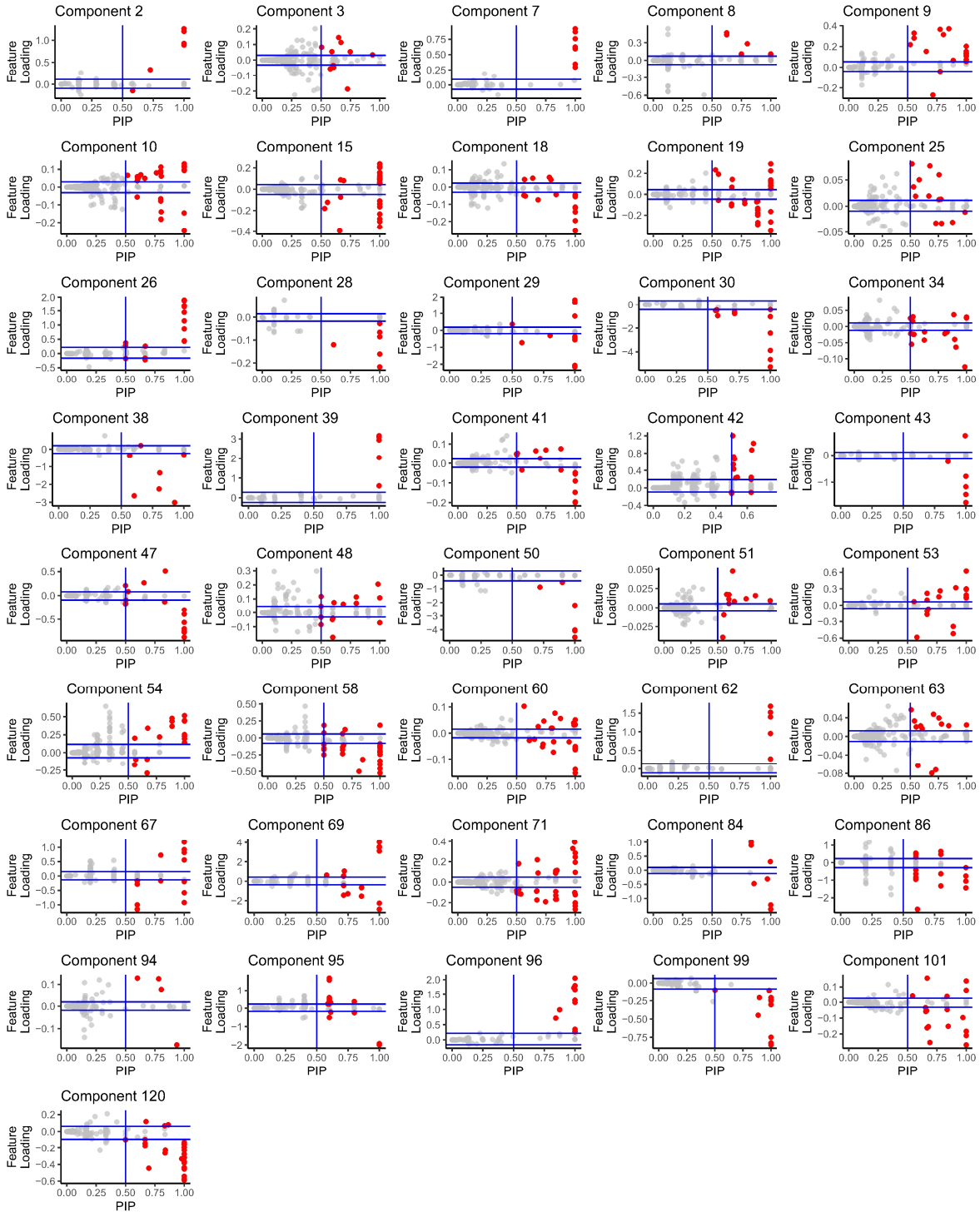


Figure 40. Feature loading scores and Posterior Inclusion Probabilities for the 41 robust BCR repertoire SDA components identified using hierarchical clustering.

BCR repertoire features shown in red have a PIP score greater than 0.5 and feature loading score greater or less than the mean ± 1 SD (blue threshold). These are interesting candidates for follow-up analysis. SDA identified both bidirectional and unidirectional components. As expected, given the use of a sparse 'spike and slab prior' which enables the shrinkage of feature scores to zero, for most components the majority of BCR repertoire features contribute very little to the component.

3.5.3.3 Correlation between SDA Components and WGCNA Modules

Repeated measures correlation (RMCORR) was used to quantify the association between BCR rWGCNA module score and SDA component sample score, to determine whether similar biological signals were captured across computational methods. This would provide additional evidence for the robustness of each approach [**Figure 41**] (Bakdash and Marusich, 2017). However, results show that the majority of SDA components were correlated with more than one rWGCNA module, suggesting overlapping biological signals. This is likely due to components containing bidirectional feature loading scores (these opposing features will have clustered into separate modules during rWGCNA analysis), the potential for features to have non-zero loading scores in multiple different components and also high-level repertoire organisation. Although, for the majority of components there were typically one or two particularly strong significant correlations with specific rWGCNA modules which are shown in **Figure 41. B**.

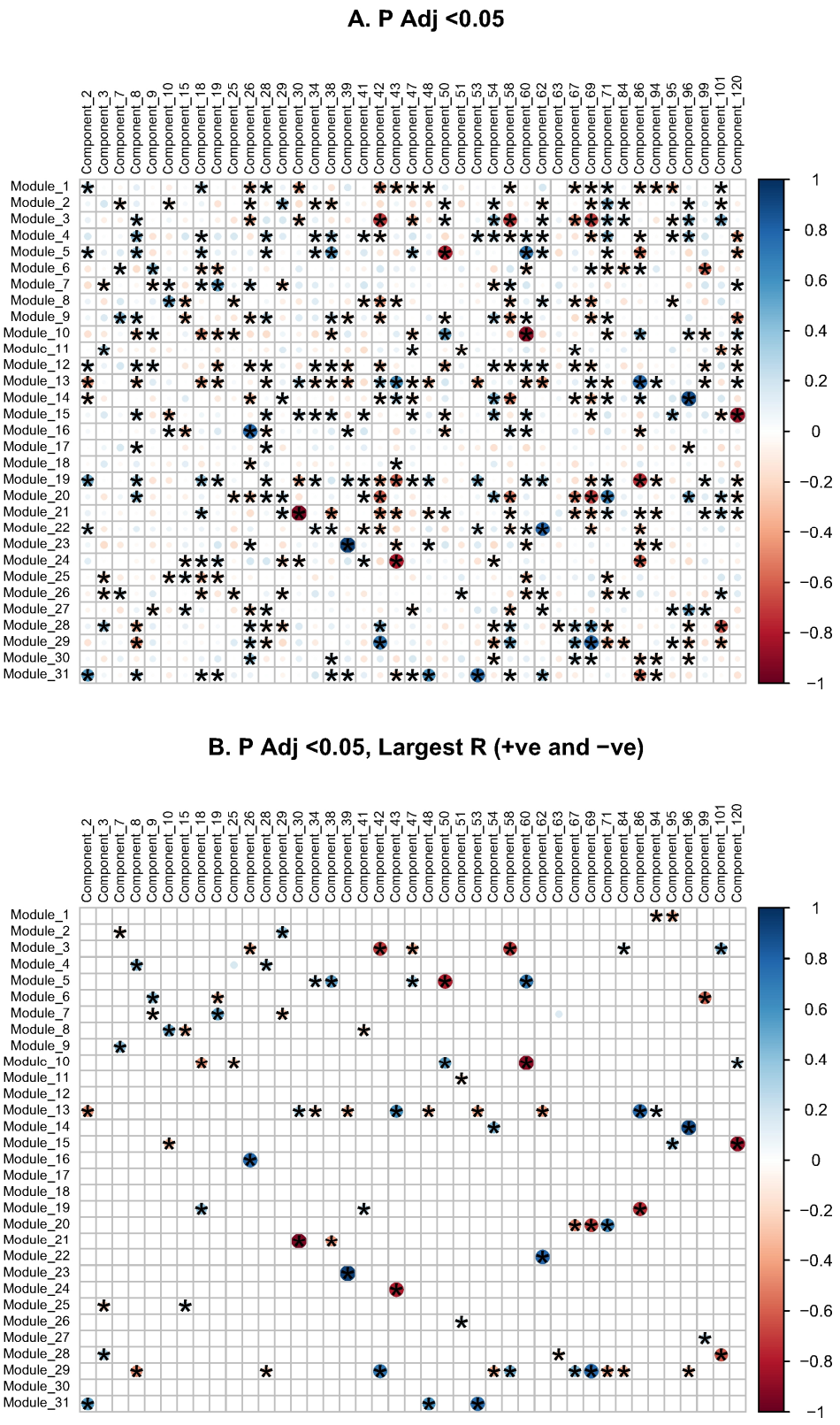


Figure 41. The correlation between BCR modules (rWGCNA) and BCR components (SDA) calculated using repeated measures correlation (RMCORR).

Significance is reported after Benjamini-Hochberg p value correction for multiple testing. **A)** All correlations between components and modules. **B)** The largest positive and negative correlation for each SDA component versus BCR rWGCNA module, to identify those that capture similar biology.

To determine whether the strongly correlated component and module pairs shown in **Figure 41.B** captured similar repertoire features, the overlap between module features and component features with PIP>0.05 and feature loading score greater or less than the mean ± 1 standard deviation (shown in red in **Figure 40**) was calculated [**Figure 42**]. The overlap for strongly negatively correlated pairs are reported in Appendix **Figure B. 2**. In some cases, good feature overlap was observed, indicative of a robust biological signal. For example, all five component 39 features overlap with module 23 (R_{rm}=0.94, 45% of features). Interestingly, there were also cases of large overlap between strongly negatively correlated WGCNA modules and components *e.g.* component 120 and module 15 (50% overlap, R_{rm}=-0.85) [**Figure B. 2**].

In the optimised rWGCNA pipeline, PC1 (module score) was orientated such that an increasing score correlates with an increasing mean 'expression' across constituent features for ease of interpretation [**Figure 36**]. Yet, the presence of substantial feature overlap between negatively correlated module and component pairs, suggests that this was not the case for SDA (increased sample score ≠ increased mean), complicating downstream analysis and interpretation. Overall, for a large number of components the overlap in features was poor. This may be due to a number of factors including the greater number of tensor components (41 versus 31), the inclusion of bidirectional loadings and all features in each component and finally, calculating the overlap with all module features rather than filtering for those with a large loading score.

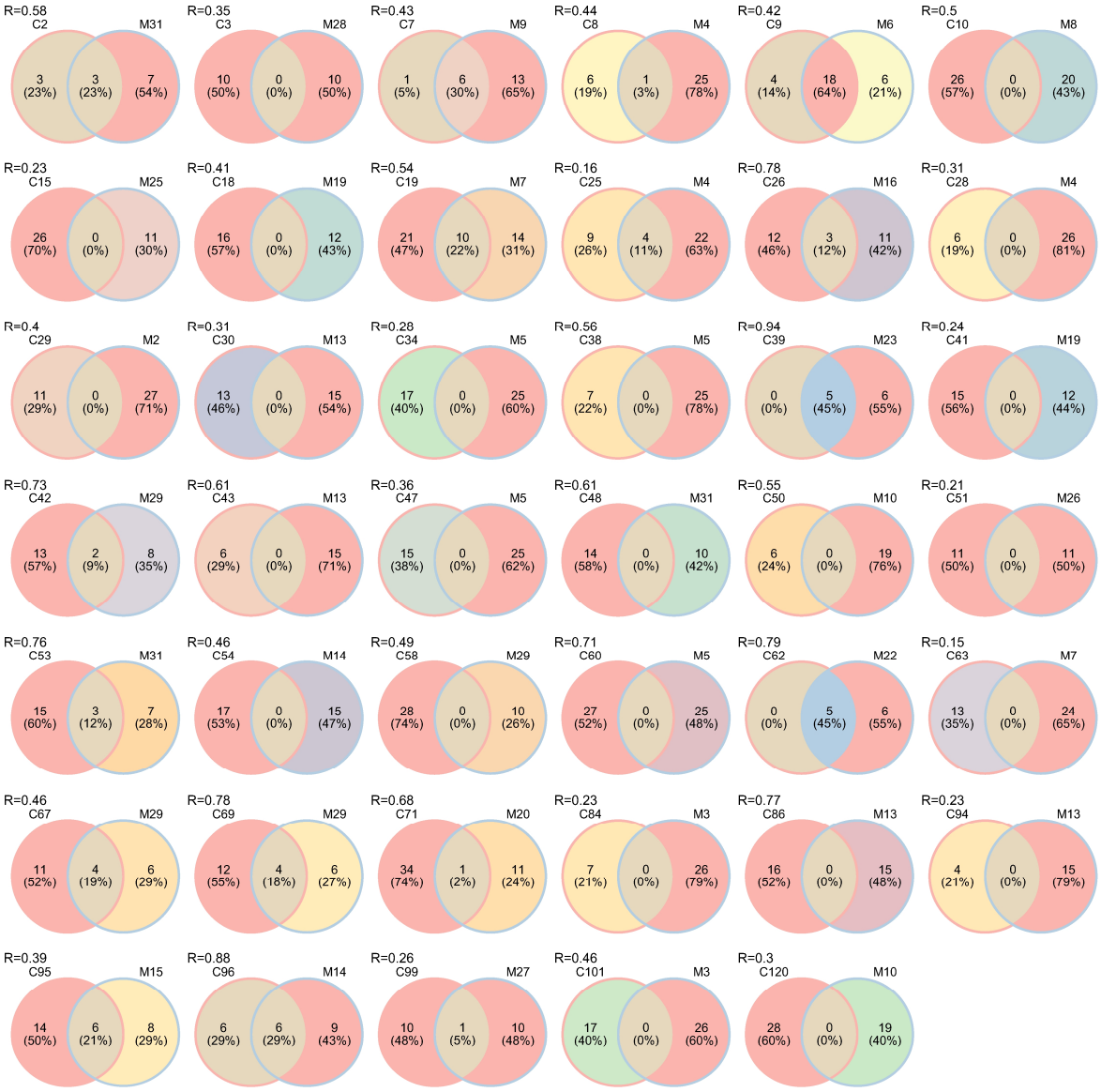


Figure 42. The number (and percentage) of overlapping BCR features between strongly positively correlated SDA components (C) and rWGCNA modules (M).

The RMCORR coefficient is shown for each component and module pair. Modules and components with a large feature overlap are likely to capture similar biological signals.

3.5.3.4 SDA Summary

In this section (**3.5.3**), I demonstrate how the machine learning approach SDA, can be used to perform dimensional reduction of BCR immune repertoire feature data. In total, 41 robust BCR IgH components capturing co-regulated repertoire features were identified, and PIP

and loading scores can be used to identify interesting candidate BCR repertoire features for further analysis from within these components.

In some cases, rWGCNA modules and SDA components captured overlapping biology, suggesting that trends in the data are robust to dimensional reduction approach. However, bidirectional versus unidirectional components, and the variation in number of features in each component/module complicated method comparisons.

3.6 Discussion

3.6.1 Comparison of Dimensional Reduction Methods

In this chapter, two different techniques were used to reduce the dimensionality of BCR repertoire data; an optimised WGCNA-based pipeline and SDA. Both methods were successful, however, each has specific advantages and limitations.

A major difference, which affects the interpretability of results, is in how methods handle anti-correlated features. During rWGCNA co-expression network generation, anti-correlated features are considered to be more distant than positively correlated features with the same absolute Pearson Correlation Coefficient. Therefore, each BCR module contains only features which co-vary in the same direction, indicated by loading scores with the same sign. The only exception to this was a small number of non-informative modules that explained low variance and included IGHV gene usage features with highly variable loadings. Unidirectional modules are particularly useful for understanding biological pathways and simplify quantitative trait analysis, as the direction of the relationship between the trait of interest and individual features within a module is clear. In contrast, SDA usually identified BCR components with bidirectional feature loadings [**Figure 40**]. This is in contrast with

the original publication by Hore et al. (2016) where typically, components identified gene networks with unidirectional gene loadings.

Bidirectional loading scores make it easy to identify features that are negatively regulated without requiring an additional analysis stage, where cross-module correlations are calculated. However, sample scores are harder to interpret as they represent the combined effect of positive and negative loadings and this consequently complicates quantitative trait analysis. The detection of bidirectional loadings in this data is likely driven by the smaller matrix and contrasting immunoglobulin isotype biology such as class-switched versus non-class switched isotypes. To simplify biological interpretation and for ease of downstream analysis I proceeded with rWGCNA modules rather than SDA components, to understand alterations to the BCR repertoire in sepsis.

3.6.2 Unbalanced Design

The unbalanced paired design of the longitudinal sepsis cohort, due to sampling truncated by death, presented unique challenges in the application of WGCNA. While the majority of BCR features were well clustered using a Pearson correlation network derived on all available time points and within modules, features showed similar trajectories across time, this approach was not perfect. There were some instances where the trajectories of constituent features within a module diverged over time [**Figure B. 1**]. This behaviour is driven by an under-representation of day 5 samples. My work expands on the methodology published by Li et al. (2018) and emphasises the importance of considering a balanced design for studies planning on utilising WGCNA methodology.

In future work, an alternative statistically correct approach for deriving the correlation network in studies with paired design could be to use repeated measures correlation

(RMCORR) to account for non-independence of samples (Bakdash and Marusich, 2017). RMCORR estimates the common regression slope or 'association shared between individuals' using ANCOVA to statistically adjust for inter-individual variability (**2.5.1**). Like the PCC the RMCORR coefficient is bounded between -1 and 1, hence it could be easily integrated into existing code. However, RMCORR can produce unstable results where study design is unbalanced and individuals are not equally represented or where the effect slope differs between individuals, as discussed by Karas (2022). Therefore, in studies such as ours, which are interested in characterising the differences between study participants that died early in disease and those that survived or died later, unbalanced design is an inherent limitation, which is not easily mitigated.

3.6.3 Consensus Module Generation

A different approach to account for non-independence of samples is to generate consensus (longitudinal) modules, as discussed in Langfelder and Horvath (2007). First co-expression networks could be generated at each time point and then those modules that are present in all networks would be used for downstream analysis. Typically, consensus module detection is applied to discrete traits such as sex or disease status, where researchers are interested in understanding the difference between groups (Langfelder and Horvath, 2008, Chai et al., 2022, Fang et al., 2021). The use of consensus modules in longitudinal analysis could result in a considerable loss of features and limit the biology which can be explored in downstream analysis. In the case of the sepsis BCR repertoire data, which has a relatively low number of observations compared to a gene expression matrix, features that diverge are easy to identify and ignore based on feature loading scores. Therefore, this additional stage may add unnecessary complexity (Langfelder and Horvath, 2008). It would also require that time is treated as a discrete rather than continuous variable, which is not an accurate

representation of dynamic network structure. Features may gradually diverge over the course of infection as individuals return to baseline health or die.

However, an alternative and interesting use for consensus module detection would be to compare the BCR repertoire networks from this sepsis cohort to those derived independently on different disease cohorts, processed using the same pipeline. By identifying consensus modules, it would be possible to determine whether there are persistent relationships between BCR repertoire features robust to sequencing methodology or disease state. This is quite likely given the overlaps between immunoglobulin isotype biology. However, modules containing IGHV or IGHJ gene usage would likely be excluded, as these are highly personalised and often disease or pathogen specific.

To expand on this project in the future, consensus modules could be provided as part of the rWGCNA code. Repertoire features would be grouped according to the consensus module feature sets, enabling direct dimensional reduction and module score generation, without the need for prior network generation. Then, cross-study comparisons could be quickly performed based on the expression of the consensus modules. This meta-analysis would provide an interesting foundation for future work, but is out of the scope of this thesis.

3.6.4 Imputation and Missing Data

Missing data can be broadly classified into three types according to the underlying mechanism, these are summarised in Kang (2013) [**Table 21**]. Understanding the origins of missing data is important for guiding the choice of imputation method.

Missing Completely at Random (MCAR)	The missing observations are a subset of all observations and there is no systematic difference between the missing and observed data.
Missing at Random (MAR):	Missing data is systematically related to the observed variables present in the data (*e.g.* sex) but is not related to the unobserved data. For example if males and females record their weight and sex, but men are less likely to record their weight, missing data will be related to the observable trait male sex.
Missing Not at Random (MNAR):	The probability of missing data is directly related to the unobserved data or information that is not present in the dataset and hence is not measured by the researcher. For example if males and females record only their weight, but men are less likely to submit their recording, missing data will be related to the unobservable trait male sex.

Table 21. The three categories of missing data.

Understanding the mechanism underlying missing BCR repertoire data is important as it can affect the choice of computational method used to perform missing data imputation.

In the BCR dataset, missing data occurred where the read depth for any individual immunoglobulin isotype did not exceed the user-specified sub-sample depth. This resulted in a pattern of missing data where all subsampled metrics for an individual isotype were missing and predominantly affected the rare isotypes (IGHE and IGHG3/4). It is possible that immunoglobulin isotype-specific read depth and corresponding summary statistics are uncorrelated. Therefore missing data would be the result of differences in total read depth (measurable) due to random factors such as inaccurate one-to-one sample pooling and human pipetting error. Under this assumption, data is classified as missing completely at random and in this scenario Random Forest missing data algorithms such as missForest perform well, particularly where features in the matrix are correlated (Tang and Ishwaran, 2017). Indeed, accurate imputation was observed when imputation method optimisation was performed using artificial missing data seeded completely at random, in the complete-cases subset of the BCR matrix (section **3.5.2.2.1**).

However, in practicality, it is possible that the repertoire (and hence the summary metrics describing the repertoire) of individuals with low isotype-specific read depths are

systematically different compared to those with high read depth and no-missing data. This could be due to some unobservable characteristic which has distinct independent effects on each isotype or alternatively influences all isotypes in a predictable manner. In the latter case, the values of more common immunoglobulin isotypes could inform on the missing values of rarer isotypes.

There is some evidence that read depth captures aspects of biology, as despite using similar quantities of input RNA, the technical replicates (from a healthy individual) which were sequenced alongside the sepsis patients, had markedly higher read depths [**Figure 20**]. Furthermore, lymphopenia in sepsis, which would be expected to affect read depth and the BCR repertoire, is associated with both 30 day and 100 day mortality (Finfer et al., 2022). Interestingly for IGHD V gene replacement frequency, all imputed values were higher than the original mean [**Figure B. 3**]. However, this was not the case for the IGHG3 V gene replacement frequency where imputed values fall around the mean. This suggests that the measures for non-missing isotypes may be informative of the rare missing isotypes.

Therefore, an alternative imputation approach which may yield more accurate results, would be to subset the matrix and perform imputation for sepsis and healthy patients separately, or stratify patients based on mortality, which acknowledges the possibility of phenotypic differences between groups of patients. However, the loss in sample size may limit any gain in accuracy. An interesting follow-up analysis would be to perform differential gene expression analysis between those patients with and without missing repertoire data using the bulk RNA sequencing gene expression. The presence of differentially expressed genes may suggest that data is not missing at random.

3.6.5 AIRR Minimum Standards and Package Development

rWGCNA was developed in R code to use the output matrices from our custom in-house pre-processing pipeline, rather than as a standalone software package. As discussed in **3.1.1,** new AIRR-seq analysis tools are encouraged to use input data in standard AIRR format where applicable, so that tools are available to the wider community. The rWGCNA code can be applied to any BCR or TCR repertoire feature matrix, generated using standard AIRR tools (*e.g.* VDJ tools, IMMCANTATION) and will perform data imputation, network generation and PCA analysis.

In future, it will be important to develop the latter half of the pre-pressing pipeline (feature generation) as a standalone R package, using IMGT summary files as input. Currently it is closely interlinked with the pre-IMGT stages of the pre-processing pipeline [**Figure 16**]. This would enable rWGCNA to be run quickly and easily on publicly available data using the same BCR repertoire features. Hence, external validation cohorts could be incorporated into larger meta-analyses, helping to address the limitations of this study. It would also facilitate consensus module detection as discussed in **3.6.2**.

3.6.6 VDJ Sequence Representation

In this chapter, I focused on the development of a novel computational approach to reduce the dimensionality of repertoire *summary* features including CDR3 length, SHM and V gene usage. To realise the full potential of AIRR-seq data, the development of computational tools that capture the nucleotide and amino acid composition of BCR and TCR sequences is required, given their importance in antigen recognition.

One solution currently employed in the literature, is to use natural language processing (NLP) to encode amino acid sequences as vectors of real numbers or 'embeddings' which

capture physical and chemical properties which contextualise the BCR/TCR within the larger repertoire (Leem et al., 2022, Ostrovsky-Berman et al., 2021). An example of this is 'immune2vec' which encodes CDR3 regions by considering three amino acids at a time as individual 'words' that incorporate the biochemical and biophysical properties of the constituent amino acids. The entire CDR3 sequence is represented as a 'sentence' comprised of these encoded words. At the 'sentence' level the CDR3 embeddings can be used to group individual CDR3s for example by IGHV gene usage and identify CDR3 sequences with similar properties but different amino acid sequences, which is important for predicting antibody specificity. At the repertoire (or 'text') level, embeddings can be used to classify individuals, using nearest neighbours approaches. A similar method was published by Kim et al. (2021).

One of the limitations of NLP methods are their reliance on a limited number of pre-determined amino acid properties. For example, immune2rep considers amino acid hydrophobicity, polarity, charge and acidity. These features may miss hidden patterns in the data. Furthermore, each subunit (word) is treated as independent, rather than contextualised within the complete receptor sequence, which is not particularly representative given the interaction between amino acids during protein folding (Leem et al., 2022).

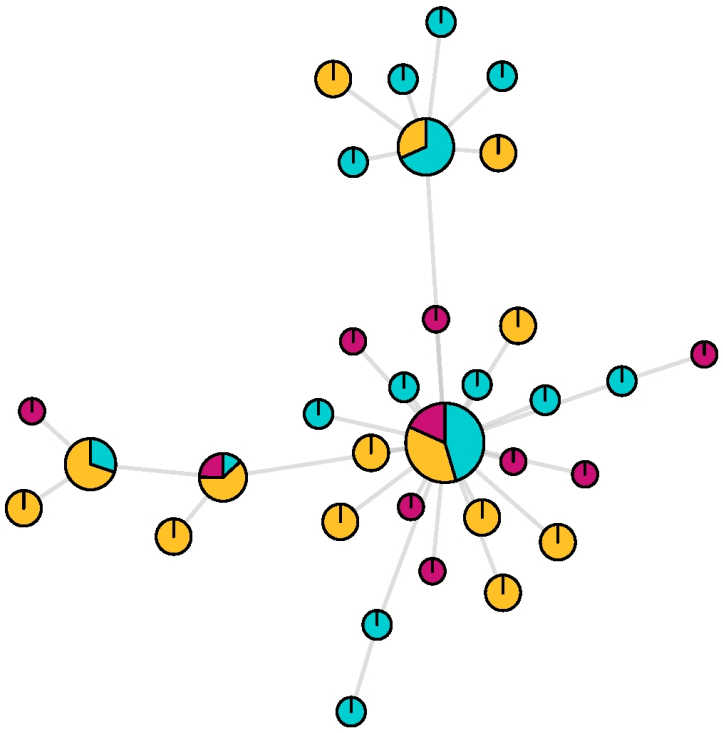
A more recent development, is the use of deep learning techniques such as in AntiBERTa published by Leem et al. (2022) which learn meaningful representations of BCR sequences based on unobserved patterns of amino acid sequences (which inform on their structure and function) often using neural networks. These BCR representations relate to B cell origin, structure and activation level.

To enrich the analysis of the sepsis BCR data presented in this thesis, it would be beneficial to apply one of these tools. It would be particularly interesting to determine whether greater separation of health and sepsis is observed when considering receptor sequences or repertoire summary characteristics, especially given the former is a much more computationally complex approach.

3.7 Conclusion

In this chapter, I have introduced rWGCNA, a novel computational pipeline for the dimensionality reduction of highly complex adaptive immune repertoire feature data, based on the principles of weighted gene correlation network analysis. I have discussed the repertoire and longitudinal data specific optimisations I performed, which will be useful to the wider research community. I have also compared the benefits, focusing on ease of interpretability, of the rWGCNA approach in comparison to the alternative machine learning approach SDA.

Chapter 4. The BCR Repertoire in Sepsis



4 The BCR Repertoire in Sepsis

This chapter summarises the statistical analysis and important biological findings for the BCR rWGCNA modules introduced in Chapter 3, generated in the context of sepsis. In the second half of the chapter, B cell longitudinal clonal dynamics are investigated using MRDARCY.

4 The BCR Repertoire in Sepsis 187

4.1 Introduction 187

4.2 Chapter 4 Aims 193

4.3 Results 195

4.4 Discussion 270

4.5 Conclusion 272

***C* Appendices: The BCR Repertoire in Sepsis 398**

4.1 Introduction

This chapter presents the results of the statistical analysis performed on the 31 BCR repertoire modules identified by the rWGCNA pipeline, introduced in **Chapter 3**. I begin with a literature review, focusing on the scientific understanding of the BCR repertoire in sepsis, followed by an exploration of the results, emphasising novel biological insights such as those associated with early mortality, disease subtypes (SRS groups) and clinical variables.

To understand regulation of the BCR repertoire within the context of sepsis, I then integrate noteworthy BCR modules, based on association with disease status or mortality, with additional experimental modalities (**4.3.2**) including:

I. Gene expression data from matched bulk RNA-seq.

II. Estimated cell fractions determined using CIBERSORTx cell deconvolution.

The latter part of the chapter employs 'MRD Assessment and Retrieval Code in Python' (MRDARCY) to characterise longitudinal B cell clonal dynamics in sepsis and health. Elements of network analysis are repurposed to gain insights into B cell clone structure in infection (**4.3.4**).

4.1.1 The BCR Repertoire in Sepsis

Research on the BCR repertoire during and immediately after sepsis has been limited, likely due to the challenges of recruiting a homogeneous patient cohort. However, this section provides a review of notable publications that have focused on the B cell response to sepsis, in either humans or animal models. These studies provide a foundation to the results presented in this chapter.

In a study conducted by Zemlin et al. (2002), which focused on understanding the development of the BCR repertoire during human ontogeny, spectratyping was used to measure the mean distribution of BCR IgH CDR3 lengths in blood samples from the foetal liver and umbilical cord from preterm and term infants. Among these participants, there were 10 premature infants which developed sepsis and their IgH-CDR3 length distribution was assessed for a period ranging from 2-12 weeks after the initial positive blood culture. Results show that the CDR3 length distribution in neonates with sepsis is consistent with oligoclonal B cell activation and expansion. Maturation of the BCR repertoire was not accelerated by systemic exposure to external bacterial antigen. However, these findings may not transfer to adult patients with sepsis due to a number of factors. Firstly, the B cell system in human infants differs significantly from that of adults and is considered premature or impaired. The neonatal BCR repertoire exhibits limited germinal centre reactions, a lack of IGHV gene mutations, delayed affinity maturation and impaired IgG/IgA class switching and decreased IgG antibodies. There is a higher rate of IgH sequence conservation, with up

to 8% of clonotypes shared between neonates. This may explain their increased susceptibility to infection (Budeus et al., 2020). Secondly, sepsis manifests differently in neonates compared to adults due to varying physiology (Wheeler et al., 2011). Notably, paediatric sepsis has a lower mortality rate (10-20%). Adult patients have a more pronounced inflammatory response, whereas the anti-inflammatory response may play a greater role in the pathophysiology of sepsis shock in children. Finally, neonatal sepsis-causing infections typically occur through transmission of pathogens from the female genitourinary system during birth and involve pathogens such as *Group B streptococcus* (GBS), *Escherichia coli*, *coagulase-negative Staphylococcus* (a commensal in adults), *Haemophilus influenza*, and *Listeria monocytogenes* (Singh et al., 2023, Molloy et al., 2020). This is in contrast to the aetiologies of community-acquired pneumonia (CAP) found frequently in septic adults in Europe (*Streptococcus pneumoniae* and respiratory viruses, mainly influenza A) (**1.1.5**). Collectively, dissimilarities in infectious etiology and physiology likely result in distinct changes to the BCR repertoire between adults and neonates with sepsis (Walden et al., 2014, Weiser et al., 2018).

Murine endotoxemia models are widely used in experimental studies of sepsis and commonly involve the injection of lipopolysaccharide (usually intraperitoneal), into mice (Lewis et al., 2016). To investigate the role of natural immunoglobulin (IgM) in the immediate response to microbial infection, Boes et al. (1998) used a mouse model of acute peritonitis, induced by ceacal ligation and puncture (CLP). Results show that mice with reduced ability to secrete IgM ($sIgM^{-/-}$ mice) have elevated 32-hour mortality (70%) compared to wild-type controls (20%). Increased susceptibility was associated with reduced levels of TNF-α, decreased neutrophil recruitment and increased bacterial load and circulating levels of pro-inflammatory cytokines and endotoxins. Low levels of circulating immunoglobulin in sepsis

are frequently reported in the literature and are predictors of high mortality (**1.2.2**) (Venet et al., 2011, Duan et al., 2020, Wilmore et al., 2018). CLP causes surgical trauma to host tissue, ischemic tissue from the ligated cecum and causes polymicrobial sepsis from faecal contamination. Therefore, standardising approaches between laboratories is challenging. Though the continuous reservoir for bacterial release is a suitable model of human abdominal infection, no source control procedure is used, making it a better model of tissue injury and untreated peritonitis than sepsis (Lewis et al., 2016).

Sjaastad et al. (2018) also employed the CLP mouse model of sepsis to explore chronic immunoparalysis. Results highlight a variety of numerical changes within the B cell compartment including a drop in B cell numbers in the blood, spleen and inguinal lymph nodes after sepsis, which returned to pre-sepsis levels by day 30, along with reduction in circulating IgM. Furthermore, CLP treated mice showed a reduced primary antibody response when subsequently immunised with Complete Freund's Adjuvant emulsion containing the MHC II-restricted peptide 2W1S56–68 at day 2. This reduction was characterised by decline in the number of cells producing IgG1, IgG2, IgG3 and IgA and a decrease in the number of antigen-specific naïve/memory and GC B cells, and plasma cells. When mice were immunised at day 30 after surgery, the reduction in these populations was less extreme, suggesting initial signs of recovery. CLP mice also showed a reduced primary antibody response following influenza A virus challenge. Collectively, these results demonstrate sepsis induced long-term detrimental alterations to the B cell compartment, increasing susceptibility to infection.

To investigate the longitudinal human plasma protein IgG1 repertoire, Bondt et al. (2021), developed a novel experimental approach coupling liquid chromatography and tandem mass spectrometry (LC-MS) which was applied to two healthy controls and eight critically ill

patients with sepsis, post-surgery. This revealed that the plasma IgG1 protein repertoire is dominated by a limited number of clones which are unique to each individual in both health and sepsis patients. Secondly, the IgG1 repertoire remained relatively stable over time, although individual clones adapted rapidly to changes in physiology, for example several clones were undetectable in plasma until the fourth time point post-sepsis (between day 3 and 61). The authors hope to develop this method further, to enable simultaneous sequencing of multiple serum immunoglobulins at the protein level.

4.1.2 Age Related Changes to the BCR Repertoire

Sepsis affects all age groups, but increasing age is a risk factor. This can be attributed in part, to immunosenescence, which is the decline in immune system function associated with normal physiological aging (He et al., 2021). Immunosenescence is characterised by the systemic and severe decline in adaptive immunity with advancing age, resulting in an increased susceptibility to infection and decreased vaccine efficacy. Corresponding alterations to the B cell compartment include a decreased number of B cells, loss of naïve cells, reduced BCR diversity, elevated levels of autoantibodies, inappropriate clonal expansions (larger and more frequent), impaired B cell selection, a shift towards the memory compartment and skewed immunoglobulin usage (Martin et al., 2015, Dunn-Walters, 2016, Tabibian-Keissar et al., 2016, He et al., 2021).

In an interesting study conducted by Arvey et al. (2020), a panel of peptides that bind serum antibodies was used in conjunction with a machine learning model to develop a score of 'immune age'. Though this score strongly correlated with chronological age, cases where immune age exceeded chronological age (accelerated immune ageing) were associated with autoimmunity, autoinflammatory disease and acute disease flares.

As summarised in He et al. (2021), there are striking similarities between age-related changes to the B cell compartment and those observed in sepsis, in particular B cell depletion. However, our understanding is limited by the small number of B cell-focused studies in sepsis. Therefore, including age as a covariate when analysing the adaptive immune repertoire in sepsis is important.

4.1.3 Sex Related Changes to the BCR Repertoire

Multiple studies have provided evidence of a sex bias in the incidence and outcome of sepsis, which coincide with sex-related differences in the immune response to infection. For example, Sakr et al. (2018) conducted an analysis of 3,902 sepsis patients admitted to ICU in Italy; female patients were significantly older and were less likely to have severe sepsis and septic shock on admission and to develop these syndromes during hospital stay. Across the cohort, mortality was similar between sexes, however, of the patients who developed severe sepsis, mortality was greater in women than in men (63.5% compared to 46.4%). In contrast, in a smaller study performed by Nasir et al. (2015), involving 97 hospital-admitted sepsis patients (not limited to ICU), men had a 70% greater mortality rate compared to females, which the authors attribute to differences in respiratory tract infection rate and IL-6 plasma levels. This is supported by a retrospective study of 261,255 critically patients which shows an increased mortality in males younger than 50 years old compared to females, although this difference was not observed in older patients (Mahmood et al., 2012, Kondo et al., 2021). Additional evidence of a sex bias is provided by a mouse model of sepsis (CLP) (Zellweger et al., 1997). A significant increase in survival was seen for pro-oestrus female mice compared to males. Interestingly, McGowan et al. (1975) reported a higher incidence of bacteraemia infections in males with sepsis compared to females, with similar case-fatality ratios.

The impact of male and female hormones on the immune system is an active area of research. Male hormones have a suppressive effect on cell-mediated immune response resulting in reduced splenocyte proliferation and decreased MHC class II expression, while female sex hormones may offer protective effects (Angele et al., 2014). Unsurprisingly then, there is evidence of sex-related changes to the BCR repertoire reported in the literature (Ricker et al., 2021). For instance, in patients with Systemic sclerosis (a multisystem autoimmune disease associated with polyclonal B cell hyper-reactivity) female patients showed greater BCR diversity than male patients (Shi et al., 2020). Similarly, a study into the human BCR repertoire in skin cutaneous melanoma (SKCM) found elevated BCR diversity in females compared to males, although there was no significant difference in BCR abundance (Selitsky et al., 2019). Jones et al. (2019) measured the total serum antibody isotype concentration in different strains of mice and showed that in all strains, antibody levels were higher in females and there was a female preference for IgG2b production. These findings were attributed to oestrogen response elements located in the enhancer elements of the immunoglobulin heavy chain, including one site, which marks a polymorphism associated with the oestrogen-augmented autoimmune disease, lupus, in humans. It is important to consider sex bias in both sepsis and BCR repertoire analysis, by incorporating sex as a covariate in statistical analysis.

4.2 Chapter 4 Aims

This chapter aims to characterise the dynamic changes to the BCR repertoire in sepsis due to community acquired pneumonia. Main objectives are:

1) To determine the most suitable statistical model for analysing the BCR repertoire rWGCNA modules, considering covariates such as age and sex (**4.1.2-4.1.3**). This will

be addressed using Akaike Information Criterion (AIC) model selection, which is a popular method for comparing the performance of non-nested statistical models as it considers both accuracy and parsimony, penalising models of high complexity (Wagenmakers and Farrell, 2004).

2) Utilising the statistical models from aim 1, to identify BCR repertoire modules that are significantly associated with:
    i) the presence of sepsis
    ii) sepsis mortality
    iii) Sepsis Response Signatures (SRS and SRSq).
3) Continuing from aim 2, to identify the major BCR repertoire features that drive module score in disease-associated modules, using PCA loading scores. Additionally, to explore the biological relevance of these modules by examining their associations with clinical data, inter-module associations, and reviewing the current literature.
4) To gain a deeper understanding of the modules from aim 2 by performing integrative analyses, examining the correlation between BCR module score and:
    a) Gene expression data from matched bulk RNA sequencing, to identify novel biomarkers associated with repertoire function and elucidate potential downstream and upstream regulatory mechanisms.
    b) Estimated cell fractions derived from bulk RNA sequencing data using CIBERSORTx.

    This aim will be addressed using quantitative trait analysis as described in Seo et al. (2016).

5) To conduct gene pathway analysis with the eXploring Genomic Relations (XGR) tool, using the set of genes significantly associated with BCR module score from aim 4.a (Fang et al., 2016).
6) To characterise persistent B cell clones using the MRDARCY tool developed by Bashford-Rogers et al. (2016). Specifically, to:
    a) Adapt aspects of network analysis to understand B cell clone structure.
    b) Describe the properties of expanding and contracting B cell clones in sepsis and in health.

4.3 Results

4.3.1 Statistical Analysis of BCR Repertoire Modules

Thirty-one BCR modules were identified using the custom rWGCNA pipeline, applied in **Chapter 3** to BCR IgH repertoire data generated for a cohort of sepsis patients and healthy volunteers. To aid in understanding the underlying biology represented by each module, the constituent BCR repertoire features for each module, annotated by immunoglobulin isotype, are summarised in **Table 22**.

Module	No. of Features	Summary	Large Loading	Immunoglobulin Isotype										
				M	D	A1	A2	G1	G2	G3	G4	E	Class Switched	ALL
1	29	Isotype Usage	N											
		Isotype Overlap	Y											
		CDR3 Length Distribution	N											
		V Gene Replacement Frequency	N											
2	27	CDR Charge	Y											
		IGHV3 V gene Usage	Y											
		Proportion of Unproductive Reads	N											
3	26	Isotype Usage	N											
		SHM	Y											
4	26	V Gene Replacement Frequency	N											
		V Gene Usage mix (V1,3,4,5,7)	N											
		Length Distribution	Y											
		Hydrophobicity Distribution	N											
		CDR3 Length	N											
5	25	Clonality	Y											
		V gene Usage V4	Y											
		Hydrophobicity	N											
		Hydrophobicity Distribution	N											
6	24	Auto-reactivity	Y											
		V gene Usage V4	N											
7	20	V Gene Usage V1	Y											
		V Gene Usage V7	N											
		Proportion Productive	N											
8	20	Isotype Usage	Y											
		SHM	N											
		Isotype Overlap	Y											
		SHM: silent to non-silent ratio	Y											
		V gene Usage V4	N											
9	19	SHM	Y											
10	19	V Gene Usage V6	Y											
		Hydrophobicity Distribution	N											
11	17	% Unmutated	N											
		V Gene Usage V3	Y											
		Hydrophobicity Distribution	N											
		Charge	N											
12	15	Isotype Overlap	N											
		V Gene Usage V4	Y											
13	15	Vertex and Cluster Rényi	Y											
		Hydrophobicity Distribution	N											
14	15	V Gene Usage V3,4	N											
		V Gene Usage V6	N											
		CDR3 Length Distribution	Y											
15	15	SHM	Y											
16	14	Clonality	Y											
		V Gene Replacement	Y											
		CDR3 Length Distribution	N											
17	14	V Gene Usage (V1,3,6,7)	N											
		Hydrophobicity Distribution	Y											

Module	No. of Features	Summary	Large Loading	Immunoglobulin Isotype										
				M	D	A1	A2	G1	G2	G3	G4	E	Class Switched	ALL
		Isotype Usage	Y											
18	13	V gene Usage V1,3	N											
		Hydrophobicity Distribution	Y											
		CDR3 Length Distribution	Y											
		Vertex and Cluster Rényi	N											
19	12	Clonality	Y											
20	12	Isotype Usage	N											
		Isotype Overlap	N											
		SHM	Y											
		SHM: CDR to framework ratio	Y											
21	11	Clonality	Y											
22	11	Clonality	Y											
23	11	Clonality	Y											
24	11	Clonality	Y											
25	11	V Gene Usage V2	Y											
26	11	V Gene Usage V5	Y											
27	11	CDR Length Distribution	Y											
		V Gene Usage V2	Y											
		Hydrophobicity Distribution	N											
		Isotype Usage	Y											
28	10	% Unmutated	Y											
29	10	Isotype Usage	Y											
		SHM: Low mutation	Y											
		SHM Distribution	N											
30	10	Clonality	Y											
31	10	Clonality	Y											

Table 22. BCR rWGCNA module summary table.

For each BCR module the constituent repertoire features are recorded and annotated by isotype where relevant. Features where the PC1 loading score was greater than expected if all features contributed equally to the eigenvector are annotated with 'Y'; these are interesting candidates for follow-up analysis.

4.3.1.1 Akaike Information Criterion Model Selection

To determine the most suitable statistical model for assessing changes in BCR module score associated with sepsis while avoiding model overfitting, model selection using the Akaike Information Criterion (AIC) was performed. Model selection was conducted independently for each BCR module.

Overfitting poses a significant concern, as it can produce inflated estimates of the statistical significance of independent variables and can generate misleading R^2 values and coefficients, increasing the risk of false positives. Overfitting may also reduce the generalisability of the model outside of the original dataset. To address this issue, AIC model selection takes into account the performance of the model, measured by the amount of explained variance in the data (BCR module score) and penalises for model complexity based on the parsimony principle, which states that the best solution is the simplest (Posada and Buckley, 2004, Portet, 2020). AIC selection is a widely-used approach to investigate the adequacy of different models, particularly where models are non-nested (Wagenmakers and Farrell, 2004).

The 'lme4' R package was used to run 21 linear mixed effect (LME) models on each BCR module. Candidate models were established using combinations of two primary variables: disease status and time point and three covariates, age and sex and the Charlson Comorbidity Index (CI). Models also included a random effect term to account for non-independence of samples [**Table 23**]. The CI is a weighted index developed in 1987 to predict the risk of death in longitudinal studies involving patients with comorbid conditions which alter the risk of mortality *e.g.* diabetes and chronic obstructive pulmonary disease (Charlson et al., 1987). The CI was assumed to be zero for healthy controls. To assess multicollinearity among the independent variables, variance inflation factors (VIFs) were computed for the complete D*TASC model. VIFs were less than 5 for all independent variables except for the interaction term. Therefore, multicollinearity is unlikely to pose a problem in downstream analysis (Gareth James, 2013).

Model	Model Formula	Notes
N	Module X ~ (1|Barcode)	Null model considering random effect of repeat sampling only.
A	Module X ~ Age + (1|Barcode)	Covariate only model.
S	Module X ~ Sex + (1|Barcode)	Covariate only model.
C	Module X ~ CI + (1|Barcode)	Covariate only model.
D*T	Module X ~ Time point*Disease +(1|Barcode)	Interaction model considering only the two main effects – time point and disease status (health or sepsis). No correction for covariates.
DT	Module X ~ Time point + Disease + (1|Barcode)	Additive model considering only the two main effects – time point and disease status (health or sepsis). No correction for covariates.
ASC	Module X ~Age + Sex + CI + (1|Barcode)	Covariate only model. Model considering only variation explained by all measured clinical covariates.
DTA	Module X ~ Time point + Disease + Age + (1|Barcode)	
DTS	Module X ~ Time point + Disease + Sex + (1|Barcode)	
DTC	Module X ~ Time point + Disease + CI + (1|Barcode)	
DTAS	Module X ~ Time point + Disease + Age + Sex + (1|Barcode)	
DTAC	Module X ~ Time point + Disease + Age + CI + (1|Barcode)	
DTSC	Module X ~ Time point + Disease + Sex + CI + (1|Barcode)	
DTASC	Module X ~ Time point + Disease + Age+ Sex + CI + (1|Barcode)	
D*TA	Module X ~ Time point*Disease + Age + (1|Barcode)	
D*TS	Module X ~ Time point*Disease + Sex + (1|Barcode)	
D*TC	Module X ~ Time point*Disease + CI + (1|Barcode)	
D*TAS	Module X ~ Time point*Disease + Age + Sex + (1|Barcode)	
D*TAC	Module X ~ Time point*Disease + Age + CI + (1|Barcode)	
D*TSC	Module X ~ Time point*Disease + Sex + CI + (1|Barcode)	
D*TASC	Module X ~ Time point*Disease + Age+ Sex + CI + (1|Barcode)	Most complex model considering the interaction between disease status and time point and adjusting for the 3 clinical covariates and sample random effect.
SE: TASC	Module X ~ Time point + Age+ Sex + CI + (1|Barcode)	Run separately for each factor of disease status (health versus sepsis)
SE: DASC	Module X ~ Disease + Age+ Sex + CI + (1|Barcode)	Run separately for each time point (day 1, 3, 5)
M: M*TASC	Module X ~ Time point*Mortality + Age+ Sex + CI + (1|Barcode)	Mortality model
S: TSqASC	Module X ~ Time point + SRSq + Age+ Sex + CI + (1|Barcode)	SRSq model

Table 23. Linear mixed effect models (LME) used in AIC model selection, simple effects tests (SE), mortality analysis (M) and SRSq (S) analysis.

Both additive (+) and interaction (*) linear mixed effects models were compared, as changes in module score across time may depend on disease status. Every model included the random effect term '(1|Barcode)' due to non-independence of samples. Covariate only models, interaction models and additive models are highlighted in yellow, green and blue respectively.

Our initial hypotheses were that the BCR repertoire would be perturbed in sepsis and that some BCR modules would exhibit time-dependent changes in sepsis but not in health, given the dynamic immune response in acute sepsis. For most modules, AICc scores (AIC corrected for small sample size) fell within a narrow range, and there was no clear best-fit model, which is commonly determined by a difference in AICc greater than 2 between two models [**Figure 43**] (Burnham and Anderson, 2002). However, models including time point and disease status generally performed better, indicating changes in BCR module score in sepsis compared to health, consistent with our hypothesis. This was particularly evident for modules 1 and 3, which show a bimodal AICc distribution. Furthermore, the optimal models for modules 3, 14, 16, and 29 included the disease status by time point interaction term (D*T). This indicates a difference in the relationship between time and module score in health versus sepsis, suggestive of a dynamic immune response in disease [**Figure C. 1**, **Table C. 1**]. However, for several modules (*e.g.* module 4), no model performed better than the null model and therefore, we were unable to explain the variance in the data using our hypotheses. Interestingly, for module 2, the optimal model included only sex, emphasising the importance of including these clinical covariates in repertoire analysis, as discussed in sections **4.1.2-4.1.3**. Due to the lack of evidence of a single optimal model in most modules and to fully address our hypotheses, analysis proceeded using the complete interaction model which included all covariates and a random effect term (D*TASC).

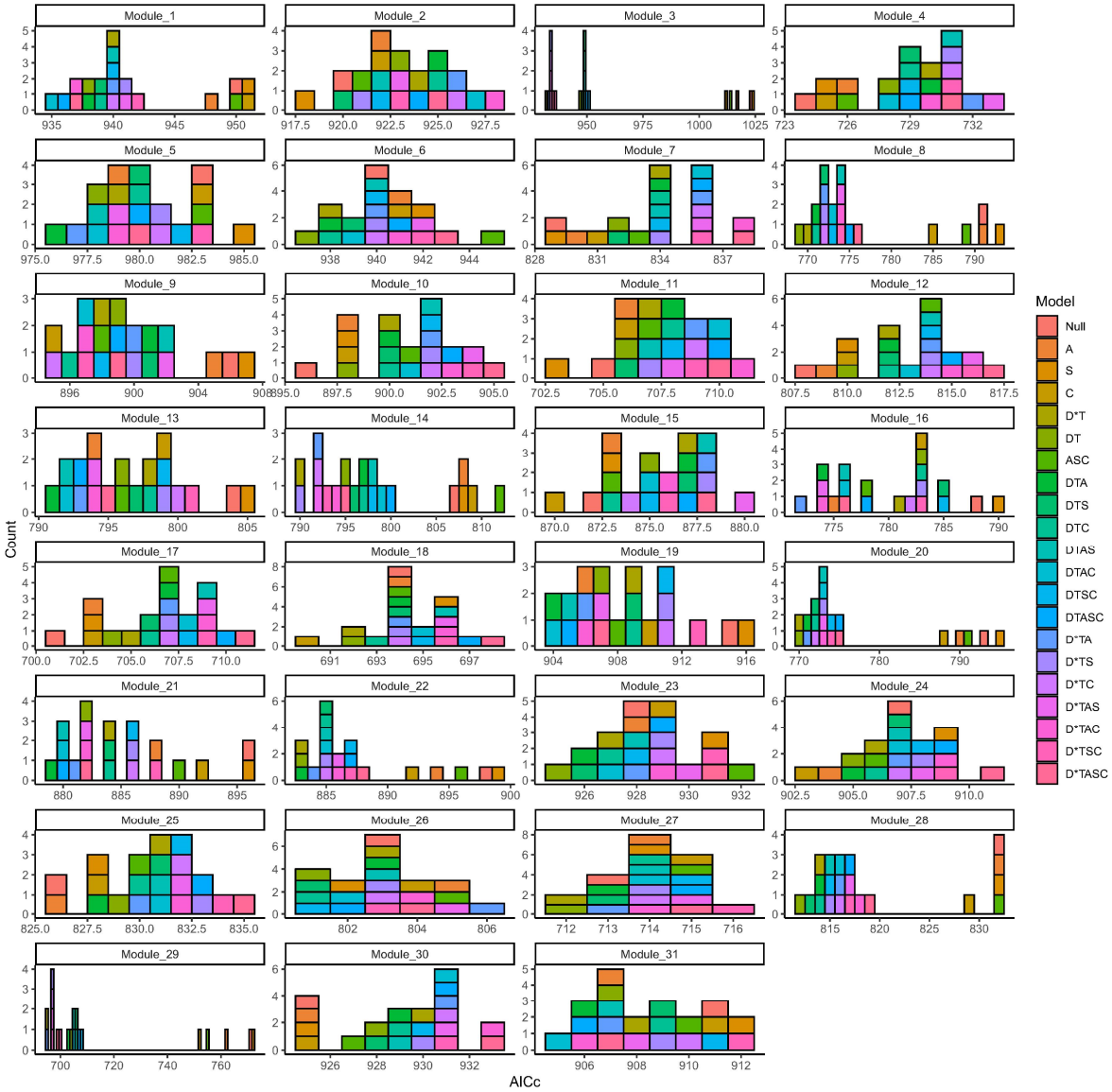


Figure 43. AIC model selection: distribution of AICc scores for each BCR module.

AICc scores for each of the 21 linear mixed effect (LME) models described in **Table 23**, applied to each BCR repertoire module separately. The AICc score is the AIC score corrected for finite sample size. The narrow distribution of AICc scores for the majority of BCR modules suggests there is no clear optimal model for downstream statistical analysis.

4.3.1.2 BCR Module Analysis: Health versus Sepsis

To identify the differences in the BCR repertoire between health and sepsis, the D*TASC LME model was applied to all modules separately [**Table 23**, **Figure 44.A**]. ANOVA was used to test whether the D*TASC model performed significantly better than either the null (model: N) or the covariate only model (model: ASC) [**Table 23**]. For modules that

displayed a significant ($p<0.05$) interaction between time point and disease status, simple effects tests were conducted. The interaction term was separated into component models to examine the effect of time in sepsis or health (model: TASC) and the effect of disease status at day 1, 3 or 5 (model: DASC) [**Figure 44.B&C**]. Partial residual plots were used to visualise the interaction between health and disease for all modules, controlling for covariates [**Figure 45**]. P values are reported in the appendices [**Table C. 2-Table C. 4**].

Seven BCR repertoire modules (1, 3, 13, 16, 19, 29 and 31) showed a significant difference in module score between health and sepsis ($p<0.05$) and one module (14) had a significant disease by time interaction but a non-significant sepsis (disease status) main effect. All eight models explained significantly more variance than the null model and 7/8 performed significantly better than a covariate only model ($p<0.05$). For module 13, the interaction model did not explain significantly more variance than the covariate only model, suggesting that variation is mostly explained by age rather than disease status.

In all four modules (3, 14, 16, and 29) analysed using simple effects tests, module score was stable in health but significantly changed over time in sepsis, indicating a dynamic adaptive immune response, clearly visualised in the partial residual plots [**Figure 45**]. Module 3 and 29 score was significantly different in sepsis compared to health at all three time points. However, module 16 was only significant at day 1, suggesting an association with those who die very early in disease. Module biology will be discussed in section **4.3.3.**

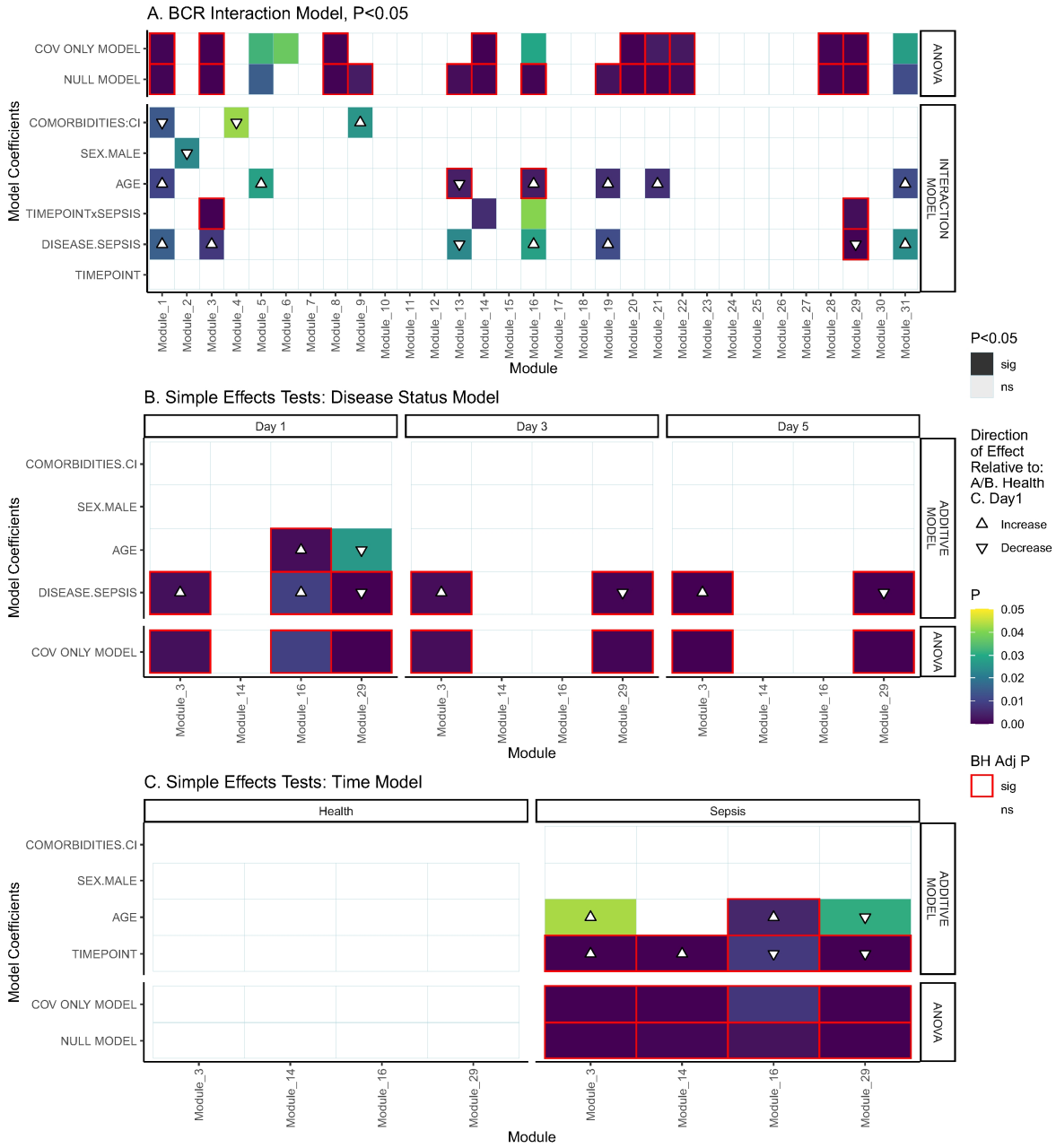


Figure 44. Linear mixed effects models comparing the effect of disease status (health or sepsis) and time on BCR module score.

A) The D*TASC model was applied to each BCR repertoire module separately and significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the interaction model explained significantly more variance than either a null model or covariate only model. **B-C)** Simple effects tests were conducted on modules with a significant interaction between time point and disease status and results are shown for: **B)** the effect of disease status (health or sepsis) at days 1, 3 and 5 and **C)** the effect of time on module score for either health or sepsis.

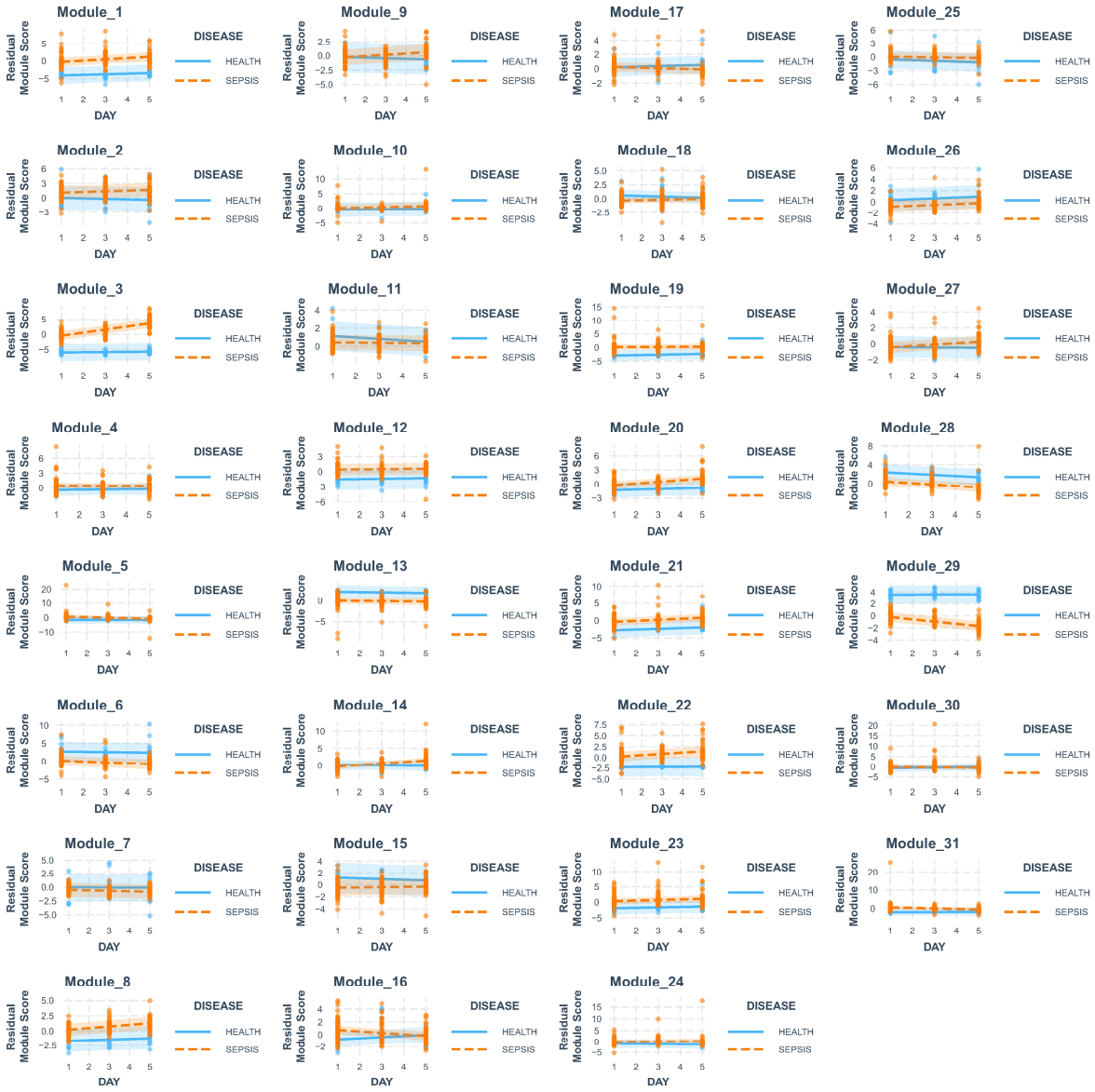


Figure 45. Partial residual plots showing the change in BCR module score in sepsis compared to health, across time, while controlling for covariates.

A partial residual plot showing the interaction between time point and disease status and their relationship with BCR module score, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (D*TASC model). 95% confidence intervals are shown.

4.3.1.3 BCR Module Analysis: Sepsis Mortality

Sepsis patients were stratified into three mortality groups: 1) non-survivors who died less than 7 days following ICU admission (early-death, ED), 2) non-survivors who died after 7 days and before 6 months post-ICU admission (late-death, LD) and 3) 6-month survivors

[**Figure 46**]. To identify differences in the BCR repertoire between sepsis survivors and non-survivors, interaction analysis was performed as in **4.3.1.2** adjusting for covariates (model: M*TASC) [**Table 23**, **Figure 47**, **Figure 48**]. Within the late-death group, the majority of individuals died less than 28 days post-ICU admission but a small number died several months after disease onset which may reflect distinct physiological processes. However, it is recommended to have at least 10 participants for every predictor level, so patients were not stratified further (Boateng and Abaye, 2019). VIF scores for the M*TASC model were all less than five, indicating that multicollinearity is unlikely to pose an issue in analysis. P values are reported in **Table C. 5-Table C. 9** of the appendices.

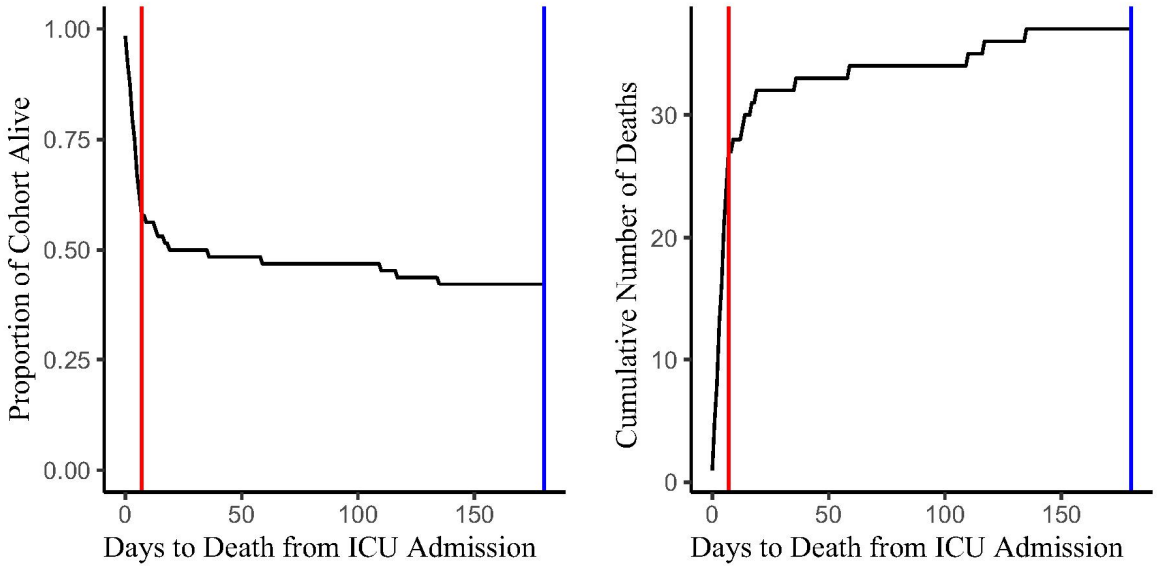


Figure 46. Stratification of sepsis patients using 7-day and 6-month mortality.

A) Proportion of cohort alive at 6 months post-ICU admission. **B)** Cumulative number of deaths. 7-day (red) and 6-month (blue) survival are marked, as these time intervals were used to categorise patients into 7-day non-survivors (early-death), 6-month non-survivors (late-death) and 6-month survivors.

Module score was significantly different in the ED group compared to survivors for 10 BCR modules (3, 4, 8, 9, 14, 16, 20, 22, 28 and 29) [**Figure 47.A**]. The LD group only showed a

significant difference in module score compared to survivors in module 9, suggesting the repertoire of patients who die late in disease is similar to survivors. However, heterogeneity in this group may mean analysis was underpowered to detect small changes.

Simple effects tests were performed on the 12 modules which had a significant interaction between time point and mortality [**Figure 47.B&C**]. Of these, 4/12 modules had significant change in module score across time in the ED group, five were significant in LD group and eight in survivors. This implies that individuals who survive acute sepsis mount a more dynamic adaptive immune response. Six modules were significantly different in ED compared to survivors at day 1, only three and two modules were significantly different at days 3 and 5 respectively. The lack of significance at later time points and absence of dynamic modules in ED may reflect sample dropout due to death; the sickest of patients have died by day 5.

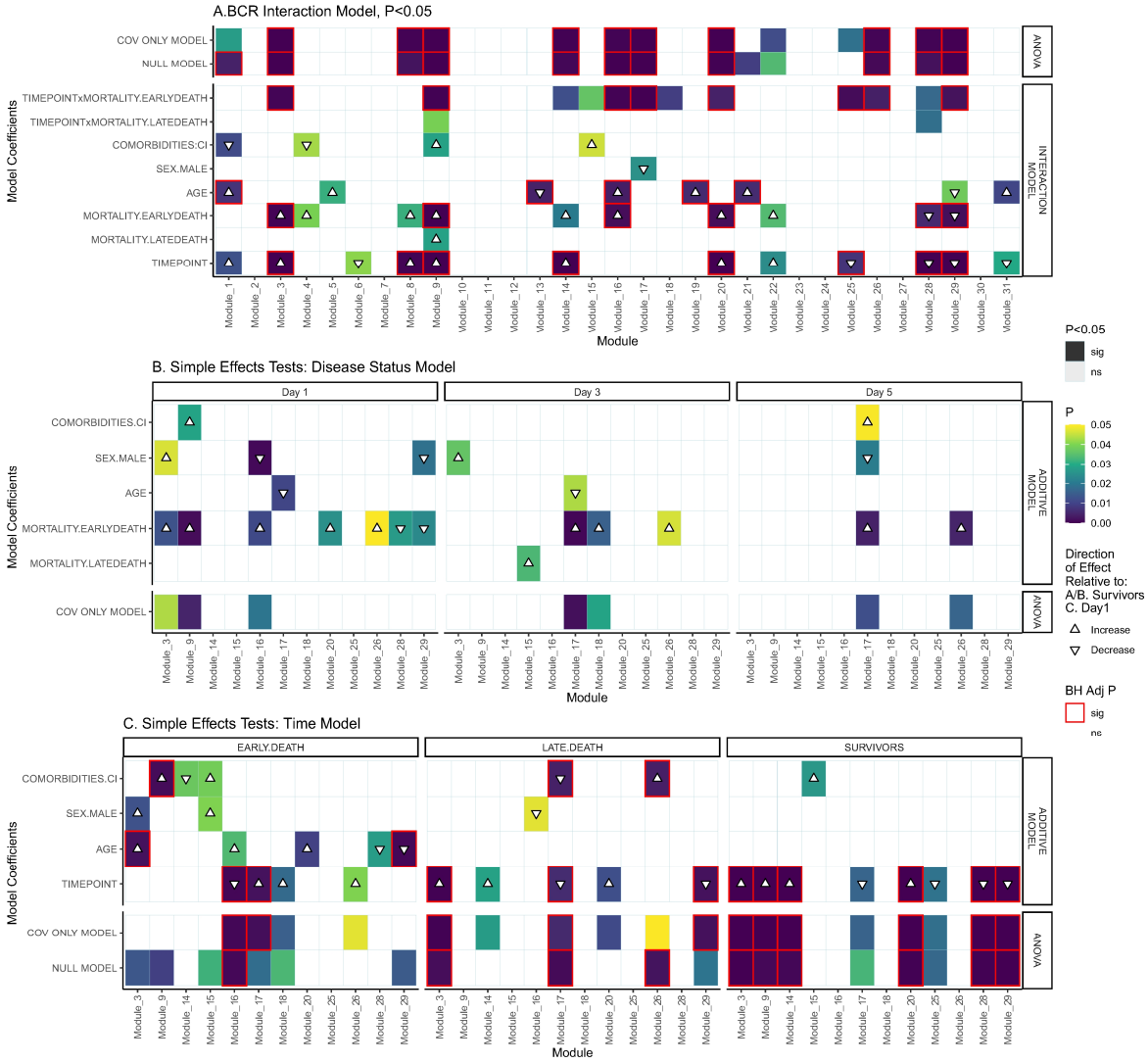


Figure 47. Linear mixed effect models investigating the effect of time and sepsis mortality group on BCR module score.

A) The M*TASC model was applied to each module separately and significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the interaction model explained significantly more variance than either a null model or covariate only model. Simple effects tests were conducted on modules with a significant interaction between time point and mortality; results are shown for: **B)** the effect of mortality at days 1, 3 and 5 and **C)** the effect of time on module score for each mortality group.



Figure 48. Partial residual plots showing the changes in BCR module score over time between sepsis mortality groups, controlling for covariates.

A partial residual plot showing the interaction between time point and sepsis mortality group and their relationship with BCR module score, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (M*TASC model). 95% confidence intervals are shown.

4.3.1.4 Inter-Module Associations: Repertoire Organisation

To understand the architecture of the BCR repertoire in sepsis, repeated measures correlation (RMCORR) was used to calculate inter-module associations (Bakdash and

Marusich, 2017). The BCR module network is shown in **Figure 49** and will serve as a basis for understanding the relationship between BCR repertoire features in section **4.3.3**.

Most modules show a degree of inter-module correlation, indicative of shared biology. This may be driven by the same BCR feature measured across different immunoglobulin isotypes. However, an added complexity is the use of similar but distinct statistical measures when characterising the repertoire. For example, the strong negative correlation between module 13 and module 19 is driven by inverse measures of clonal diversification and clonal expansion. Modules 23 (IGHA2 clonal expansion), 16 (IGHG clonal expansion and V gene replacement), 30 (IGHG3 clonal expansion) and 18 (CDR3 length and hydrophobicity distribution) are isolated, suggesting these modules capture unique aspects of BCR repertoire biology. Therefore, cluster resolution is appropriate.

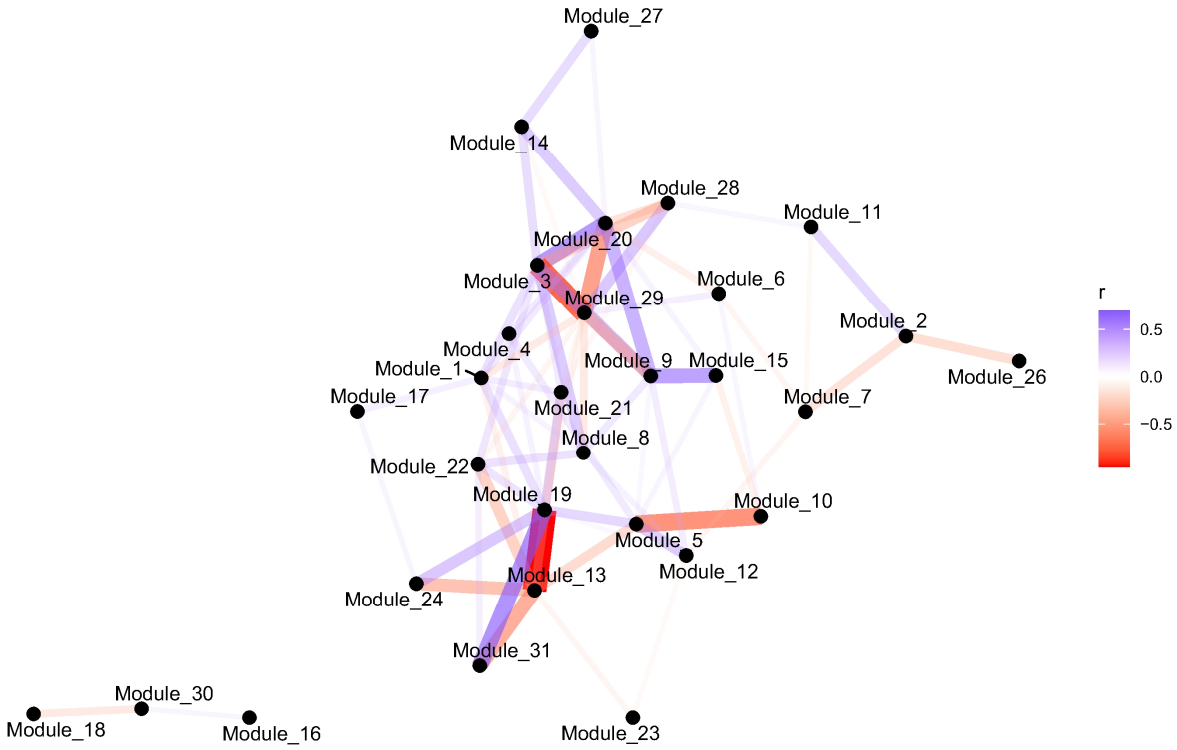


Figure 49. BCR module interaction network.

All pairwise correlations between rWGCNA BCR modules were calculated using RMCORR to understand high-level repertoire organisation. Raw p values were corrected for multiple testing using the Benjamini-Hochberg method. Only inter-module correlations with $R_{rm}>0.3$ and $p_{adj}<0.05$ are shown.

4.3.1.5 BCR Module Analysis: Association with Clinical Variables

To identify associations between sepsis clinical traits (recorded in the eCRF) and BCR module score, two analyses were conducted. Firstly, for clinical variables measured once per ICU admission, linear regression (LR) using day 1 sepsis samples was performed for each module and clinical variable separately. The clinical trait was used as the independent variable and module score was the dependent variable. The model also included age, sex and CI [**Figure 50.A**]. P values are reported in **Table C. 10**. Secondly, for clinical variables measured longitudinally, a LME model using a random effect term for sample ID and adjusting for covariates was run on each clinical variable and module, using all samples [**Figure 50.B**]. P values are reported in **Table C. 11**.

Module 14 was the sole module positively associated with disease severity, measured using the total Simplified Acute Physiology score (SAPS) and total Acute Physiology and Chronic Health Evaluation II (APACHE II) score. This trend persisted when the total Sequential Organ Failure Assessment (SOFA) score was considered at each sampling time point. A corresponding association with decreased lymphocyte count and blood pressure, and increased urea levels was observed [**Figure 50**]. Module 14 was also associated with the number of sepsis specific therapies administered, further evidence of increased disease severity. Mortality interaction analysis suggested a link between module 14 and the ED group, but this association did not endure when using simple effects tests [**Figure 47**, **Figure 48**]. Module 14 captures features including IGHV3-15 usage as well as CDR3 skewness and kurtosis [**Table 22**].

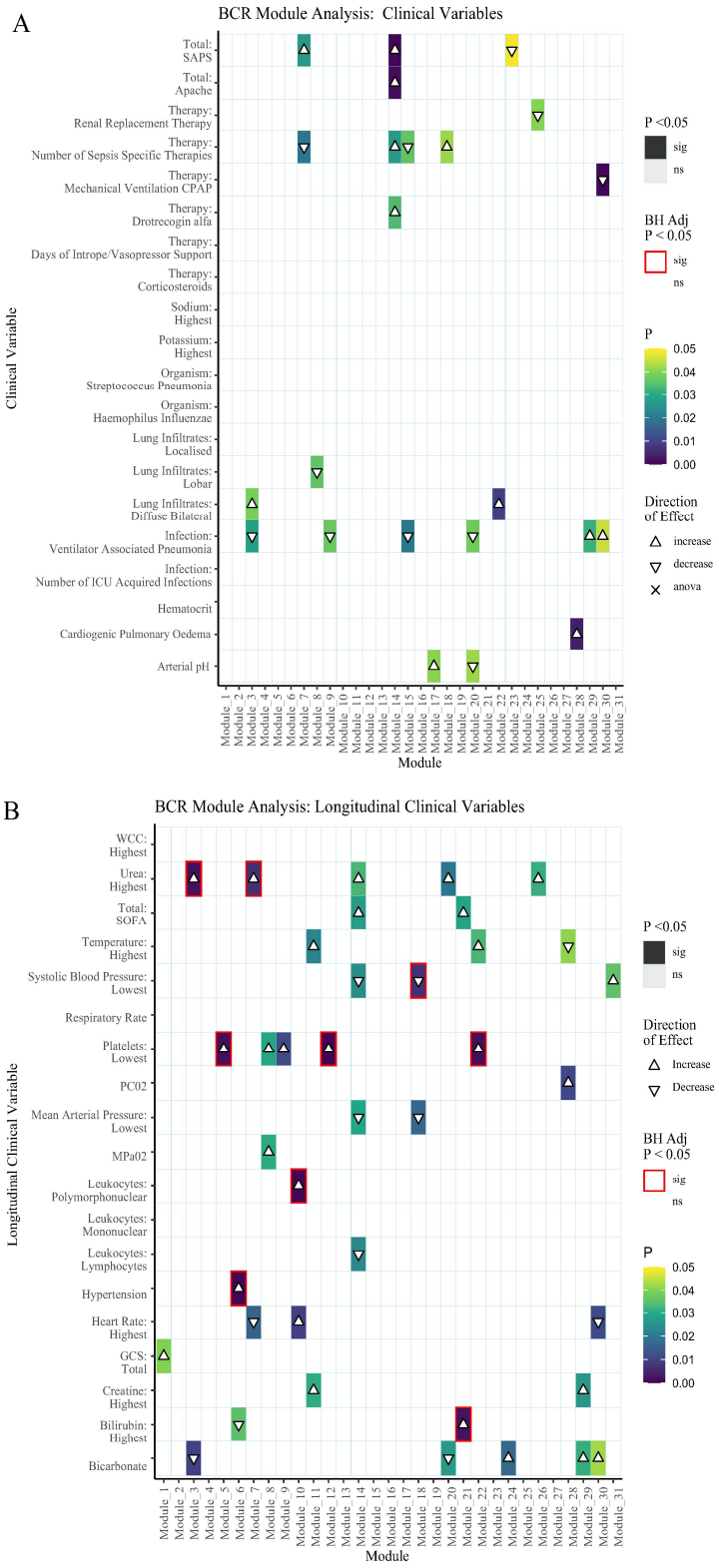


Figure 50. The effect of clinical variables on BCR module score in sepsis patients.

The effect of **A)** clinical variables measured once per ICU admission on day 1 BCR module score and **B)** longitudinal clinical variables on BCR module score (all time points). For binary clinical variables *e.g.* yes/no for presence of lung infiltrates, the direction of effect is reported relative to those without. Significant model effect terms ($p<0.05$) are shown. The BH multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red.

4.3.2 Integrative Analysis

The subsequent sections discuss the results of integrative analyses which combined the BCR rWGCNA modules and matched gene expression data from bulk RNA-seq to further understand module biology (Cano-Gamez et al., 2022, Mi et al., 2022). Out of the 205 samples initially chosen for this study, 180 had matched BCR and bulk RNA-seq after QC.

4.3.2.1 BCR Module Analysis: Association with Sepsis Response Signatures

Sepsis response signature (SRS1/2/3) and SRSq scores were assigned for samples with bulk RNA sequencing gene expression using the 'SepstratifieR' R package (Davenport et al., 2016, Cano-Gamez et al., 2022). Additive LME models (model TSqASC) were used to identify an association between BCR module score and Sepsis Response Signature quantitative (SRSq) score **[Table 23]**. Healthy controls were excluded from this analysis.

Only one BCR module (26) had a significant association with SRSq score **[Figure 51**, **Figure C. 2**, **Table C. 12]**. SRSq score is derived from whole-blood leukocyte transcriptomics, which may explain the lack of associations. Neutrophils are the most common blood leukocyte and previous research has shown that expansion of immature neutrophils in SRS1 is the likely cellular immunological basis for the SRS endotypes (Kwok et al., 2023). Around 20-40% of circulating white blood cells are lymphocytes, of these B cells comprise only ~10% (Morbach et al., 2010, Blumenreich, 1990). Therefore, the contribution of B cells to a whole blood derived transcriptomic signature will be limited. The identification of SRS-independent BCR mortality signatures in section **4.3.1.3 [Figure 47**, **Figure 48]** may facilitate improved stratification of sepsis patients based on adaptive immune response, provided appropriate biomarkers are identified.

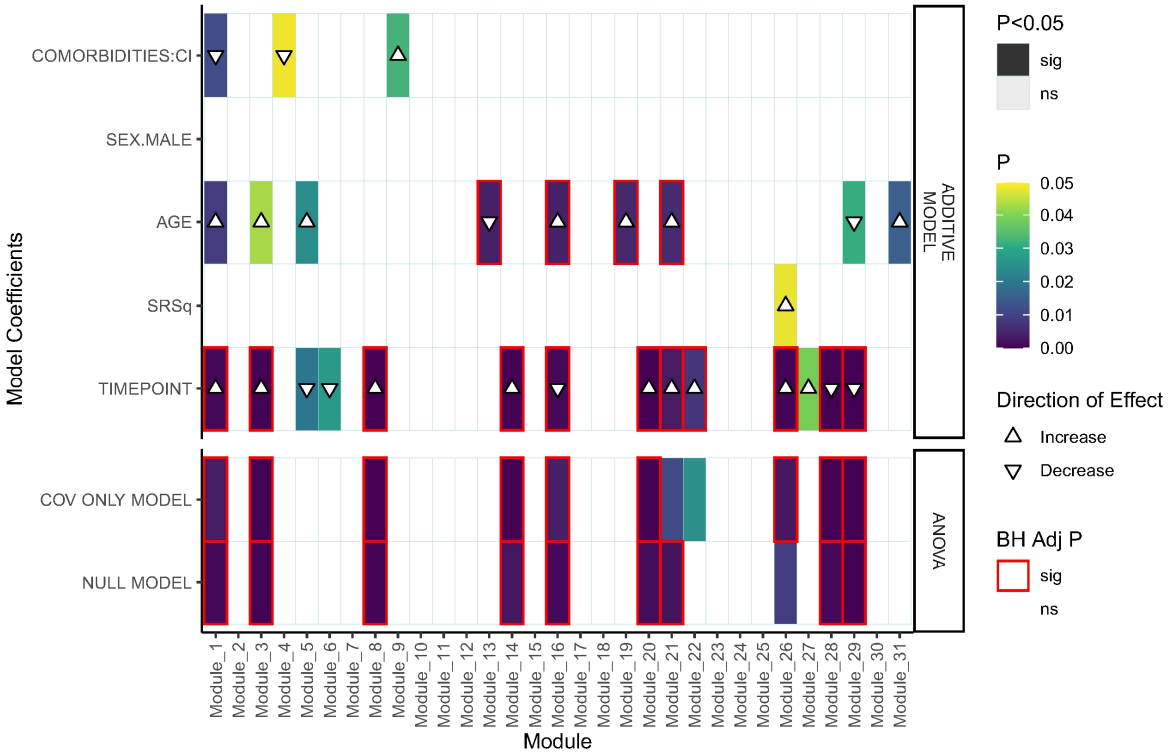


Figure 51. Linear mixed effects models exploring the effect of SRSq score on BCR module score (TSqASC).

Significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the model performed significantly better than either a null model or covariate only model.

4.3.2.2 CIBERSORTx Cell Deconvolution

4.3.2.2.1 Dynamic Changes in the Fraction of Peripheral Blood Cell Types during Acute Sepsis

CIBERSORTx was used to estimate whole-blood leukocyte cell fractions for samples with matched bulk RNA-seq (**2.4.1.5**). To understand dynamic changes in cell populations of interest (lymphocytes) in sepsis, estimated cell fraction was modelled as the dependent variable in LME models, as in **4.3.1.2 [Figure 52**, **Figure 53**, **Table C. 13-Table C. 15]**.

Results show a significant decrease in the estimated cell fraction of circulating lymphocytes (B cells, $CD4^+$ and $CD8^+$ T cells) in sepsis compared to health. This is consistent with the literature which has shown that lymphopenia is maintained 28 days after sepsis (Monserrat et al., 2013, Jensen et al., 2018). This is also consistent with the reduced number of BCR reads obtained from sepsis patients compared to in health.

A reduction in B cell fraction was observed across all time points, meaning B cell lymphopenia preceded ICU admission and did not recover to baseline health by day 5 post-ICU admission. There was no significant difference in plasmablast fraction between health and disease. This may reflect delayed activation of the adaptive immune system, as the number of plasmablasts was significantly increased in sepsis at day 5 when each time point was tested separately ($p<0.05$). Plasmablasts represent a small proportion of the total cell fraction therefore estimated count will be heavily influenced by sequencing depth.

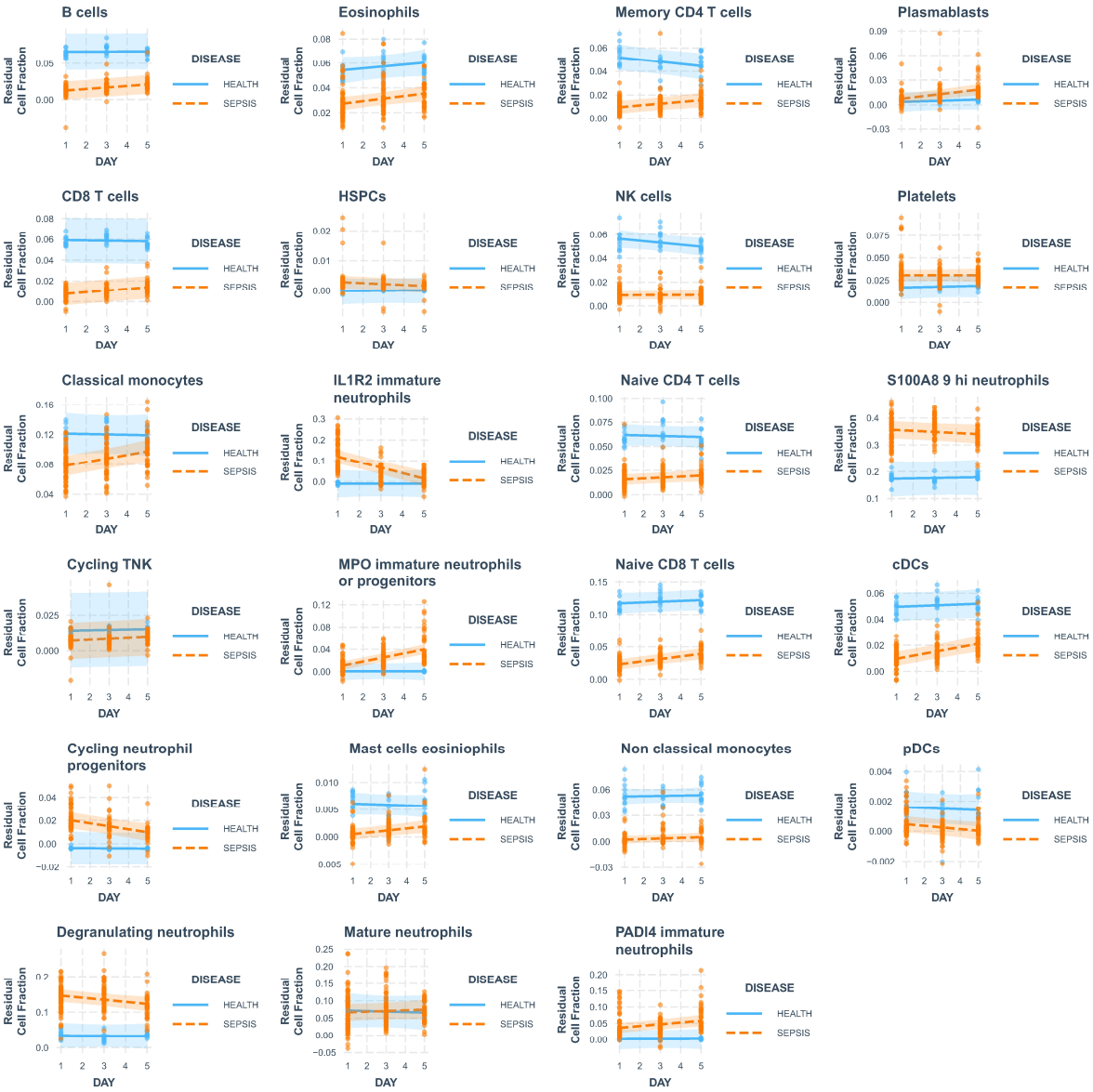


Figure 52. Partial residual plots showing the dynamic changes in CIBERSORTx estimated cell fraction in health and sepsis, controlling for covariates.

A partial residual plot showing the interaction between time point and disease status and their relationship with CIBERSORTx estimated cell fractions, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (D*TASC model). 95% confidence intervals are shown.

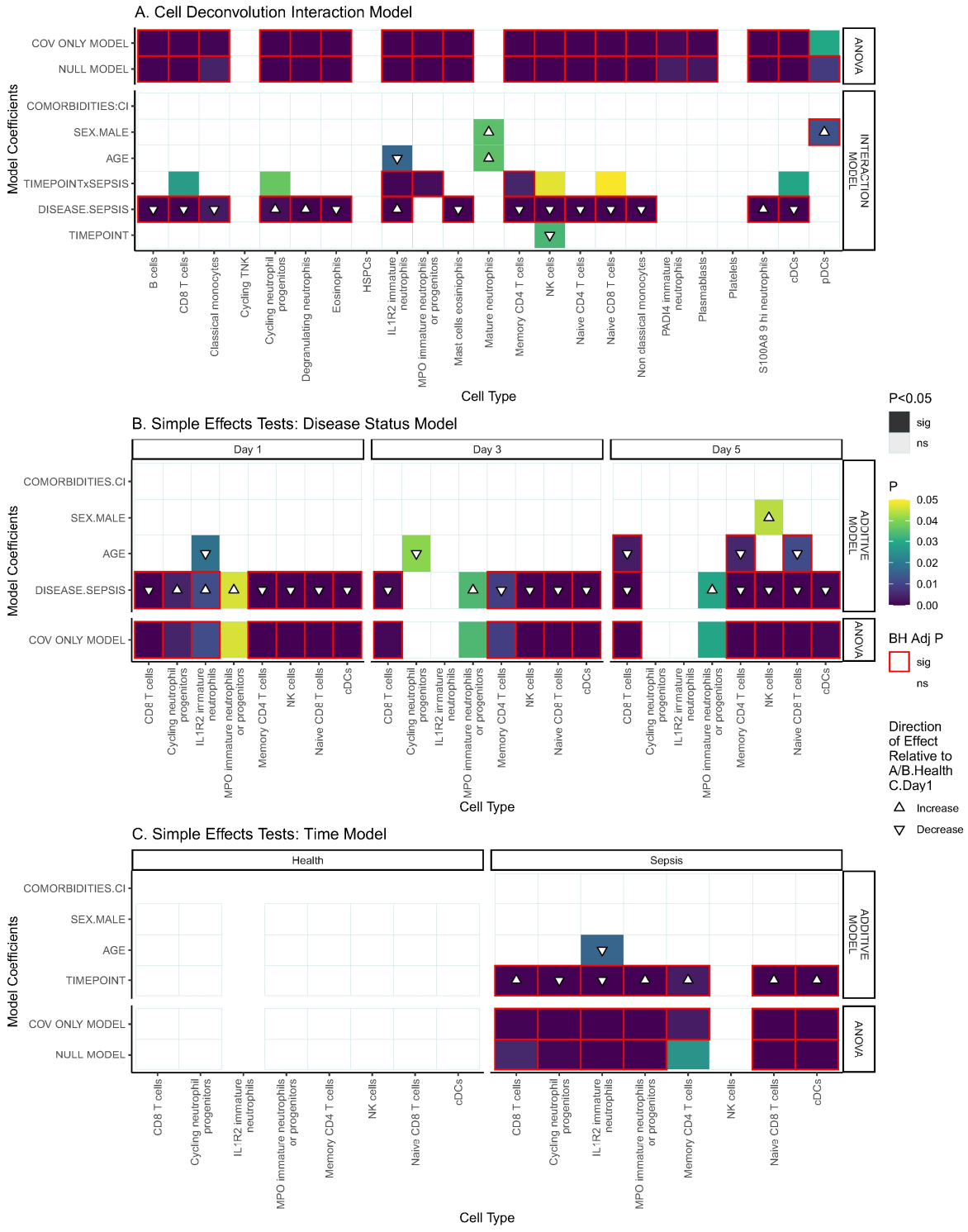


Figure 53. Linear mixed effects models comparing the effect of disease status and time on CIBERSORTx estimated cell fraction.

Significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the model performed significantly better than either a null model or covariate only model. For those models with a significant time point by disease status interaction follow-up simple effects tests were performed investigating the effect of **B)** disease status at each time point and **C)** the effect of time in health or sepsis.

4.3.2.2.2 BCR Module Analysis: Association with Peripheral Blood Cell Fraction

Quantitative trait analysis, as described in Seo et al. (2016), was used to assess the relationship between BCR module score and CIBERSORTx estimated cell fractions. LME models were employed for each module and cell type combination. Two models were run: (1) cell proportion as the independent variable to determine whether expansion of certain populations affected BCR repertoire module score [**Figure 54**] and (2) module score as the dependent variable to determine whether BCR module score influenced the fraction size of different cell types [**Figure C. 3**]. Both models were used as the direction of causality is uncertain, and x and y variables are non-interchangeable in linear regression analysis. The models included disease status, time point, and a random effect term for sample ID.

Significant relationships were identified using a minimum effect size (ES) of the mean ± 1 standard deviation and significance of $p<0.05$. Reassuringly, the majority of significant associations between BCR modules and cell types were for B cell and plasmablast fractions. Further discussion of these results can be found in section **4.3.3**.

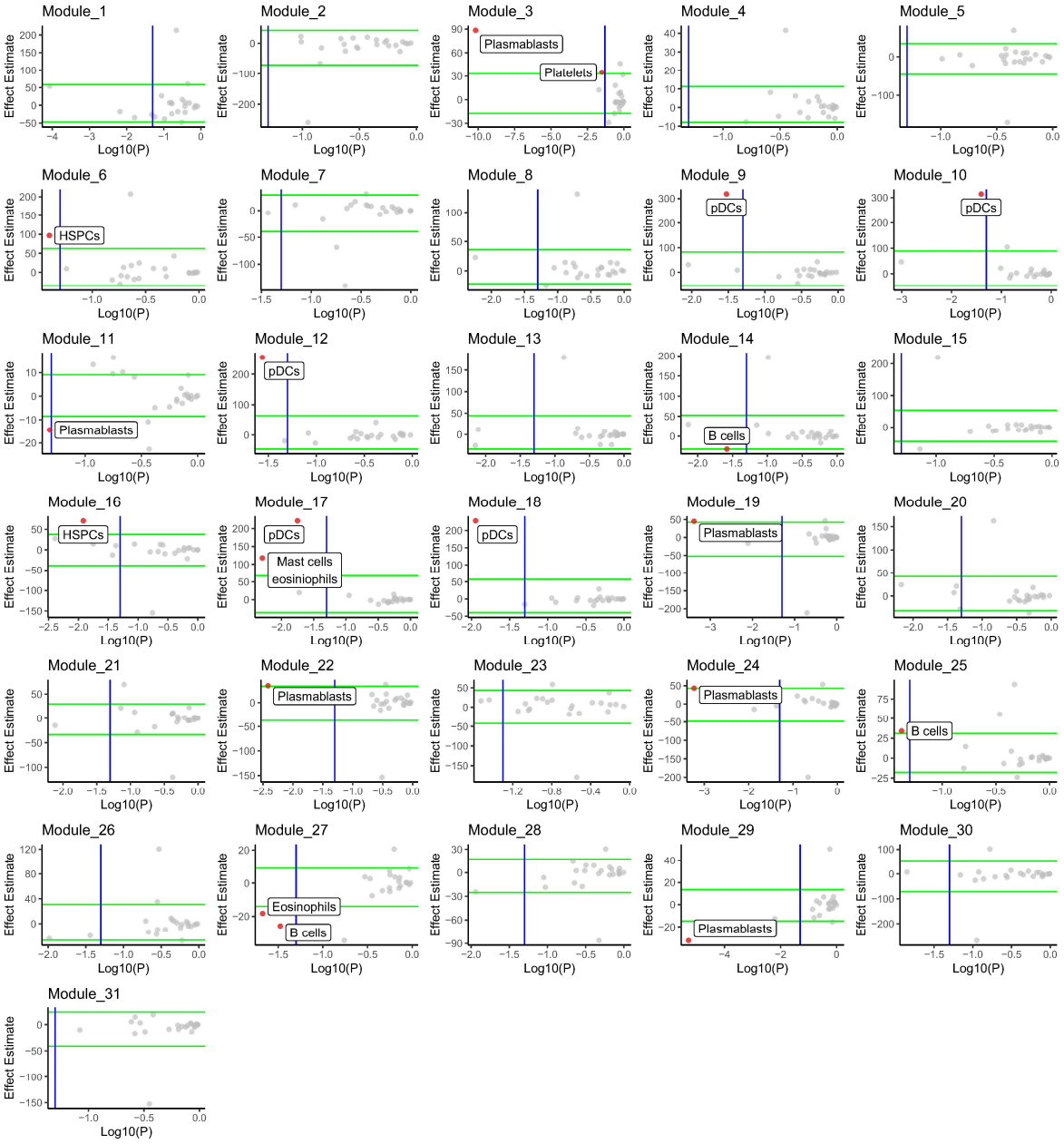


Figure 54. Association between CIBERSORTx estimated cell fractions and BCR module score using cell fraction as the independent variable and module score as the dependent variable.

Estimated effect sizes on module score for each cell fraction and their corresponding p values are plotted. Log10(p=0.05) is shown in blue and the mean ± 1 SD is shown in green. Cell types which meet these thresholds are annotated.

4.3.2.3 Immunoglobulin Isotype Gene Expression

As discussed in **2.4.3**, the normalised percentage of the repertoire explained by each immunoglobulin isotype, as output by the in-house processing pipeline, cannot be used to accurately assess absolute changes in isotype expression. Therefore, immunoglobulin isotype log-transformed gene expression counts were modelled using the D*TASC model. All isotypes except IGHE had a significant time point by disease status interaction [**Figure 55, Figure 56, Table C. 16-Table C. 18**]. Therefore, simple effects tests were used to understand the dynamic changes in sepsis.

For all isotypes except IGHE, there was a significant increase in gene expression during sepsis disease course. IGHM and IGHD expression was significantly lower than health at days 1 and 3 but not at day 5. This indicates return to baseline health in individuals that survived to day 5 post-ICU admission and depletion in IGHM and IGHD prior to ICU admission. IGHG1, IGHA1 and IGHA2 were significantly elevated at time point 5 compared to health suggestive of clonal expansion and increased class-switched antibody production.

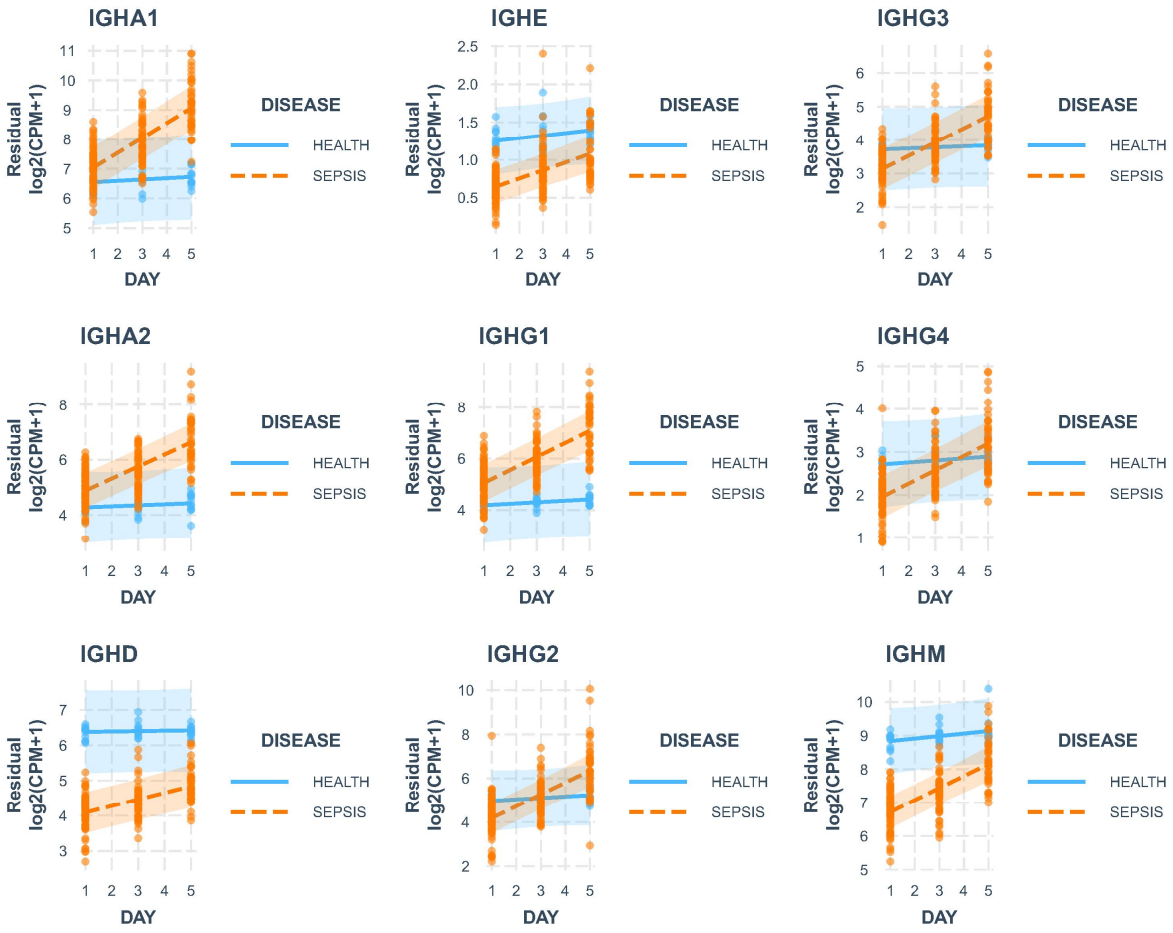


Figure 55. Partial residual plots showing the dynamic changes in immunoglobulin isotype gene expression (bulk RNA sequencing) during acute sepsis, controlling for covariates.

A partial residual plot showing the interaction between time point and disease status and their relationship with immunoglobulin isotype log-transformed gene expression from bulk RNA sequencing, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (D*TASC model). 95% confidence intervals are shown.

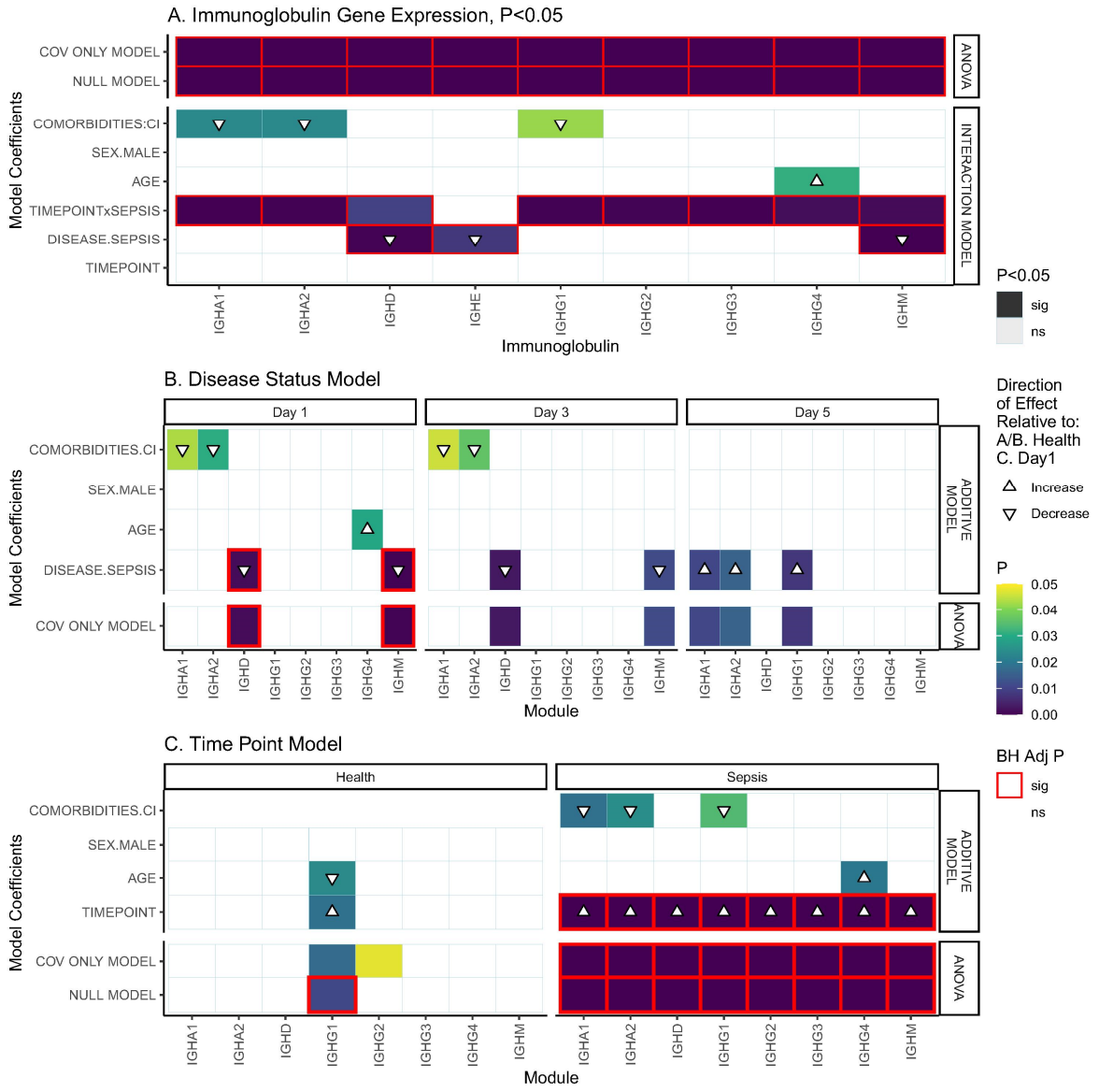


Figure 56. Linear mixed effects models investigating the effect of disease status and time on immunoglobulin isotype gene expression (bulk RNA-seq).

Significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the model performed significantly better than either a null model or covariate only model. For those models with a significant time point by disease status interaction follow-up simple effects tests were performed investigating the effect of **B)** disease status at each time point and **C)** the effect of time in health or sepsis.

4.3.3 Interpreting BCR Repertoire Module Biology in the Context of Sepsis

4.3.3.1 Module 1

4.3.3.1.1 Elevated CSR in Sepsis

BCR module 1 score was significantly elevated in sepsis and positively correlated with age. There was no statistically significant association with sepsis mortality. Drivers of module 1 (determined by PC1 loading scores) are the isotype overlap between different class-switched and non-class-switched isotypes and the percentage of the repertoire comprised of IgG1 [**Figure 57**]. Therefore, module 1 can be broadly summarised as class-switch recombination (CSR) in response to antigen exposure. Significant elevation in sepsis indicates diversification of antibody effector function in response to infection.

CSR occurs predominantly in germinal centres and gives rise to long-lived plasma and memory B cells. In agreement, module 1 score was associated with a significant increase in plasmablast fraction [**Figure C. 3**]. CSR from IgM to IgG is important for neutralising viruses systemically and IgA antibodies are critical for immunity at mucosal sites of entry. Extensive CSR to IgG and IgA classes has been documented in response to respiratory infections caused by SARS-CoV-2 (Nielsen et al., 2020b), influenza virus (Mathew et al., 2021) and *Streptococcus pneumonia* (Barker et al., 2021, Zhou et al., 2004). Therefore, elevated CSR in patients with sepsis due to CAP is consistent with the response seen for common sepsis-causing pathogens.

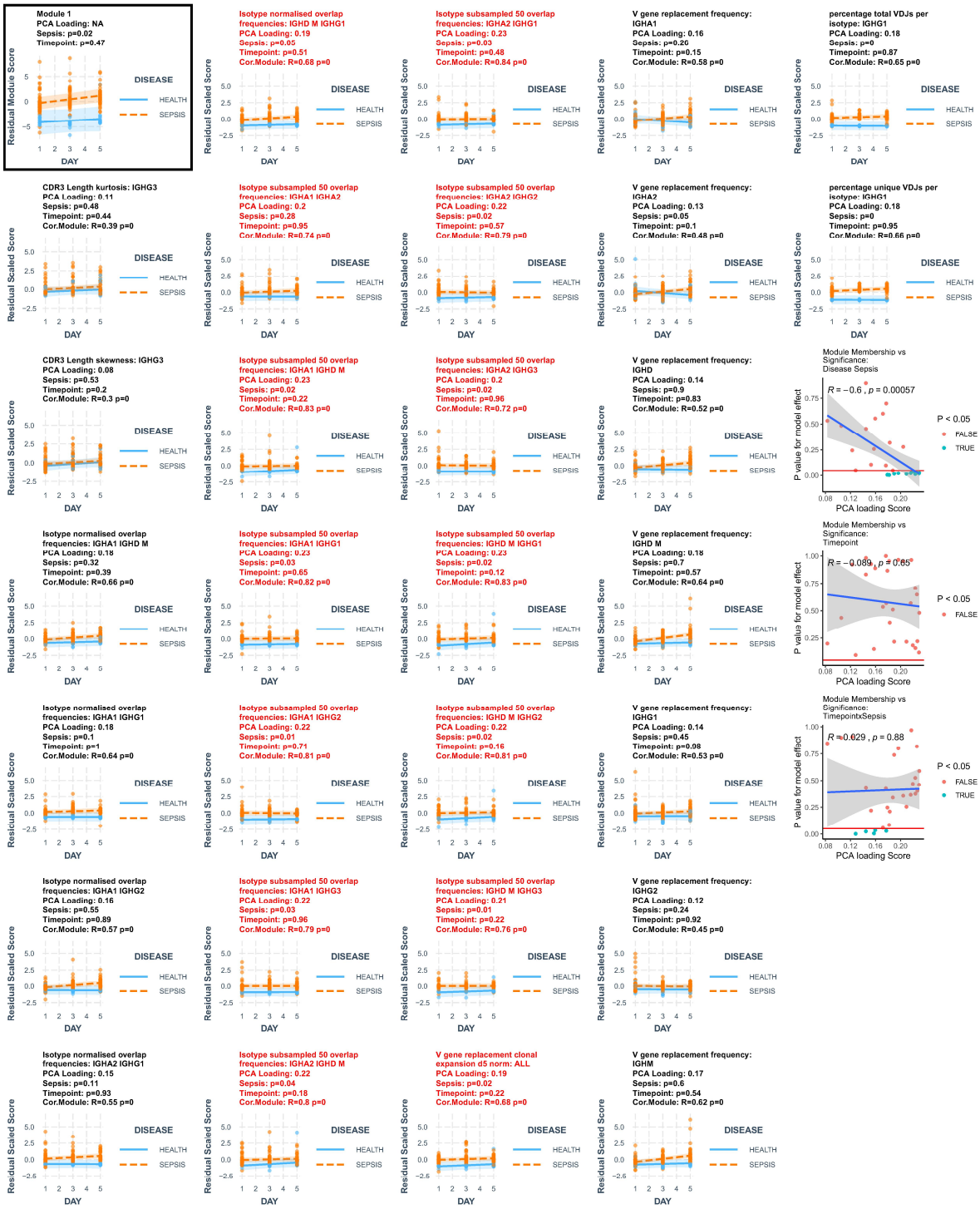


Figure 57. BCR module 1 summary.

Top left shows the partial residual plot for BCR module 1 using the D*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 1 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms. 95% confidence intervals are shown. Not all features are well correlated with the module eigenvector, and in these cases the difference between sepsis and health may not be significant.

4.3.3.1.2 Elevated IgA1-to-IgA2 CSR in Sepsis

CSR occurs by an intra-chromosomal deletion recombination event resulting in inherent directionality, due to the arrangement of immunoglobulin constant regions proximal to IGHM on the IgH locus [**Figure 6**]. No CSR is possible after IGHA2 (the most distal constant region). CSR frequency inference networks were generated using the subsampled isotype overlap frequencies, of which the majority were captured by module 1 [**Figure 58** and **Figure 59**]. Notably, in those who died less than 7 days after ICU admission (early-death) a significant increase in CSR frequency was observed across multiple pairs of isotypes from as early as day 1, compared to health. For sepsis survivors increased CSR frequency compared to health was only seen at the later time-points [**Figure 59**].

Across sepsis mortality groups the most frequent CSR event was consistently IGHA1→IGHA2 and to a lesser extent IGHG1/G3→IGHG2. Mucosal IgA is important for protection against respiratory infections via direct neutralisation or prevention of binding, but does not cause inflammation due to an inability to fix and activate the complement cascade (Cerutti, 2008) [**Table 2**]. IGHA2 antibodies are more abundant in mucosal secretions and have a reduced susceptibility to bacterial proteases compared to IGHA1 (Schroeder and Cavacini, 2010). Given the focus on patients with sepsis due to CAP, the high frequency of CSR from IGHA1→IGHA2 likely reflects a beneficial response for maintaining passive protection of respiratory mucosa during respiratory infection, via increasing antibody diversity.

Sequential IGHA1→IGHA2 CSR can be triggered upon receipt of T cell independent signals from bacteria-activated epithelial cells and DCs and is dependent on A proliferation-inducing ligand (APRIL). In the colonic lamina IgA2 non-conventional CSR has been shown

to occur in the absence of functional $CD4^+$ T cells, CD40 signals and germinal centres from mutated lamina propria IgA1 B cells arriving from mucosal follicles, perhaps as a means to maximise diversity of the IgA2 repertoire at the site of infection (Cerutti, 2008). This mechanism has also been reported in tonsillar crypts (Xu et al., 2007). Given the observation that CSR from non-IGHA1 isotypes to IGHA2 is much lower in comparison to CSR from IGHA1, elevated IGHA1→IGHA2 CSR in sepsis could be explained by a non-germinal centre CSR model involving regulation via respiratory epithelium (Xu et al., 2007). The IGHA2 BCR sequences may be detected in the blood due to leaky endothelial barriers.

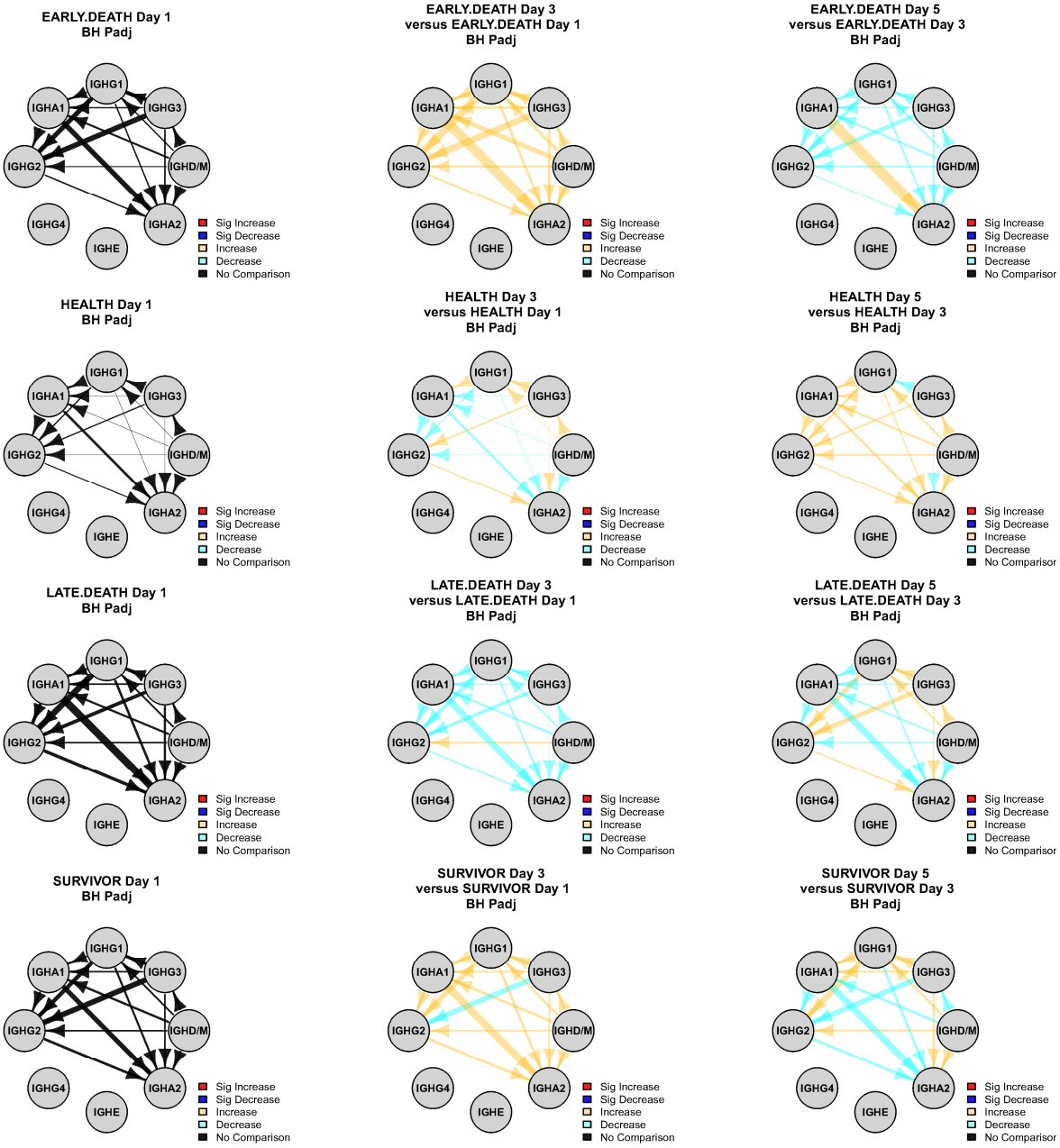


Figure 58. CSR networks investigating the effect of time on CSR frequency in health and sepsis mortality groups.

The frequency of BCR sharing between each isotype is shown by edge width, with the corresponding direction of class-switch recombination shown by an arrow, which is determined by the organisation of the IgH loci. Edges are coloured according to whether sharing is significantly or non-significantly increased or decreased relative to the same switch for the previous time-point. Significance (p_{adj} <0.05) was determined using a Wilcoxon test and BH correction for multiple testing. The lack of edges connected to IGHG4 and IGHE is due to high levels of missing data.

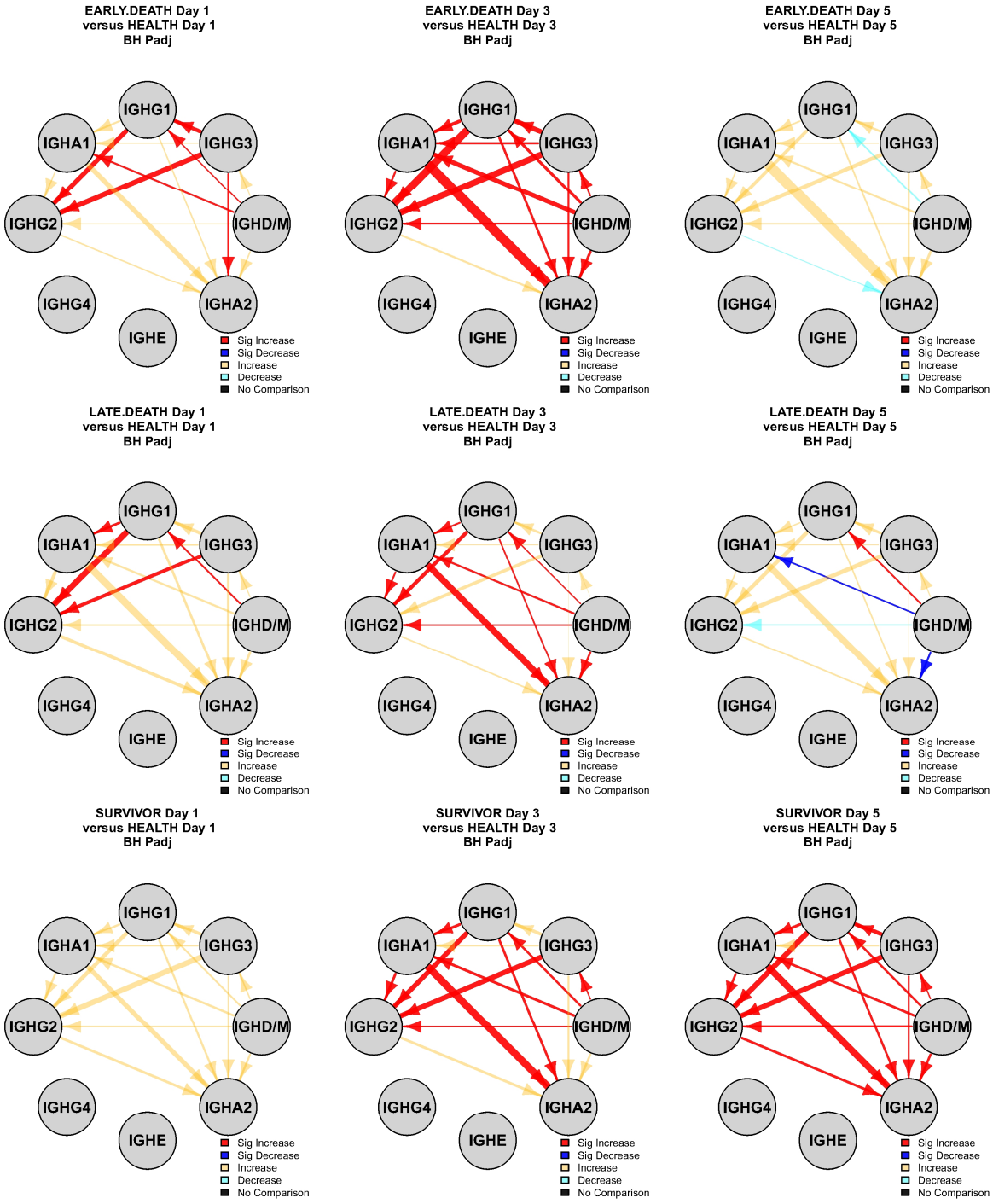


Figure 59. CSR networks comparing sepsis mortality groups to health.

The frequency of BCR sharing between each isotype is shown by edge width, with the corresponding direction of class-switch recombination shown by an arrow, which is determined by the organisation of the IgH loci. Edges are coloured according to whether sharing is significantly or non-significantly increased or decreased relative to the same switch at the same time point for healthy individuals. Significance ($p_{adj} < 0.05$) was determined using a Wilcoxon test and BH correction for multiple testing. As indicated by the LME analysis, there is significant increase in CSR in health compared to sepsis. The lack of edges connected to IGHG4 and IGHE is due to high levels of missing data.

4.3.3.1.3 Increased IGHG1 Usage is Associated with Platelet Fraction

IgG1 is secreted in response to protein antigen and causes activation of the classical complement pathway via C1q binding and bacterial clearance via opsonisation and phagocytosis [**Table 2**]. Consistent with module 1 score, analysis of bulk RNA-seq, revealed a significant increase in IGHG1 expression over time in sepsis [**Figure 55**, **Figure 56**]. IgG1 is recognised by stimulatory FcγRs (FcγRI/IIA/IIC/IIIA/IIB); binding to FcγRIIA on monocytes, macrophages, neutrophils, dendritic cells and platelets is associated with superior immune response to streptococcal infection (McKay et al., 2015). However, no association between module 1 and *Streptococcus pneumoniae* infection in the clinical data was detected [**Figure 50**].

FcγRIIA is the only Fc receptor on human platelets and IgG1 binding triggers platelet activation and aggregation (Zhi et al., 2015). Interestingly, module 1 score was associated with a significant increase in the estimated platelet fraction [**Figure C. 3**]. Module 1 was also associated with a decrease in the proportion of monocytes, perhaps due to platelet-monocyte interaction, which can propagate monocyte differentiation into tissue-resident macrophages (Kral et al., 2016). IgG1 binding to Fcγ receptors promotes neutrophil activation via tyrosine kinase pathways, though no significant association was observed with estimated neutrophil fraction [**Figure 54**] (Shankar-Hari et al., 2012).

4.3.3.1.4 Association with Non−Integrin Membrane−ECM Interactions

To identify genes significantly associated with module 1 score, quantitative trait analysis was performed. LME models were run independently on each gene from the bulk RNA sequencing data, using gene expression as the independent variable and module score as the

dependent variable [**Figure 60**]. Covariates including disease status, time point and a random effect term were included in the model. For each gene, the estimated effect size (ES) on module 1 score and corresponding p value was calculated. Genes were categorised as either adaptive immune receptor (AIR) repertoire genes, defined as those genes also captured in the targeted repertoire analysis (*e.g.* V, D, J and constant genes from BCR and TCR chains) or non-AIR genes. Pathway enrichment analysis was performed using 'XGR' and REACTOME annotations. Only non-AIR genes with a Benjamini-Hochberg adjusted p value less than 0.05 and ES greater or less than the mean ± 1 standard deviation (SD) were input into pathway enrichment analysis.

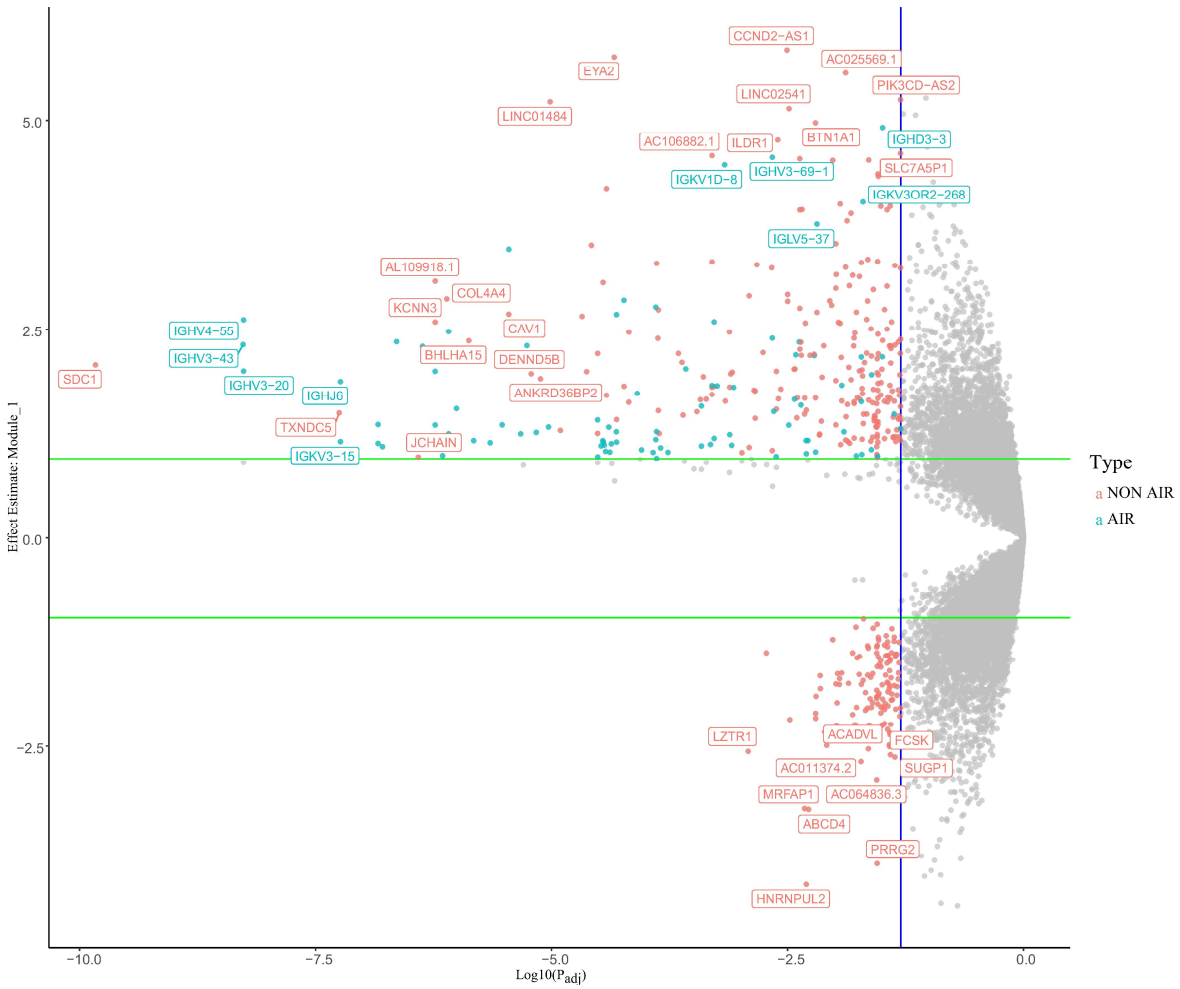


Figure 60. Quantitative trait analysis identified 451 genes significantly associated with BCR module 1 score.

Linear mixed effects models were used to identify genes significantly associated with module 1 score, using bulk RNA sequencing gene expression data. P values were corrected for multiple testing using the Benjamini-Hochberg method. The log10(p_{adj}=0.05) threshold is shown by the blue line and the mean effect size ± 1 standard deviation are shown in green. Genes are annotated as either adaptive immune repertoire genes (AIR) or non-AIR genes.

The most significant module 1 associated non-AIR genes were: J chain which is important for IgA polymerisation (p_{adj}=6.99e-08, ES=1.00), Syndecan-1 (*SDC1*, also known as CD138) which is a plasma cell marker and involved in cell proliferation, migration and cell-matrix interactions (p_{adj}=1.91e-11, ES=2.13) and Thioredoxin domain containing 5 (*TXNDC5*, p_{adj}=1.32e-08, ES=1.54) (Khodadadi et al., 2019, Castro and Flajnik, 2014).

TXNDC5 is a member of the protein disulphide isomerase family and it promotes the formation and rearrangement of native disulphide bonds ensuring proper protein folding in the endoplasmic reticulum (Hung et al., 2022, Pengo et al., 2013). Interestingly, *TXNDC5* is a marker for seroconversion (where the level of antigen-specific antibodies rises to a detectable level) and is hyper-expressed in B cells after COVID-19 vaccination (Bando et al., 2023, Huang et al., 2021). The strong association with module 1 score in sepsis provides further support for use as a novel biomarker of seroconversion, as patient samples were recruited pre-COVID-19 pandemic. During plasma cell differentiation, the size of the endoplasmic reticulum is expanded to support production of large volumes of antibodies and this coincides with enhancement of folding capacity. This suggests a role for *TXNDC5* in immunoglobulin disulphide bond formation (Pengo et al., 2013). Non-AIR genes with the largest positive ES on module 1 include angulin-2 (*ILDR1*, p_{adj}=0.002, ES=4.84). *ILDR1* has been shown to distinguish cells at the late stage of B cell differentiation (plasmablasts and plasma cells) and it is part of the immunoglobulin superfamily, mediating adhesive interactions (Strittmatter-Keller et al., 2018).

Pathway enrichment analysis identified a single significant module 1 associated pathway after multiple testing correction (non−integrin membrane−ECM interactions, $p_{adj}<0.05$) and 10 additional pathways with $p<0.05$ [**Figure 61**]. The non−integrin membrane−ECM interactions term included the genes: COL10A1 (p_{adj}=0.037, ES=2.29), COL4A3 (p_{adj}=0.017, ES=2.16), COL4A4 (p_{adj}=4.011e^{-7}, ES=2.93), DDR2 (p_{adj}=0.010, ES=3.01), SDC1 (p_{adj}=1.91e^{-11}, ES= 2.13). Cell-ECM interaction can regulate cell function including controlling the mobilisation of growth or differentiation factors and modulating cell proliferation (Rosso et al., 2004). Rapid proliferation is a hallmark of B cells undergoing GC reactions and plasmablasts. Thus, the identification of the solute carrier (SLC)-mediated

transport pathway likely represents the dramatic increase in nutrient uptake and metabolism required during B cell activation, proliferation and differentiation (Song et al., 2020).

Although not significant after BH multiple testing correction, the detection of the double-strand break repair pathway including the genes *DNA2* (p_{adj}=0.017, ES=1.76), *EXO1* (p_{adj}=0.0097, ES=1.46), *EYA2* (p_{adj}=3.51e^{-5}, ES=5.82), *PARP2* (p_{adj}=0.025, ES=1.61), *POLL* (p_{adj}=0.046, ES=-1.68) and *UBA52* (p_{adj}=0.037, ES=-1.49) is particularly interesting. CSR is initiated by AID-induced DNA breaks and aberrant processing can result in decreased CSR. Breaks are ligated by proteins from the classical-non-homologous end-joining (C-NHEJ) and alternative-NHEJ pathways. DNA damage in cells undergoing CSR promotes PARP2 expression, which suppresses deleterious IgH/c-myc (Myc proto-oncogene) translocations which occur when AID-induced DNA breaks are aberrantly resolved in B lineage lymphomas (Robert et al., 2009). EXO1 (exonuclease-1) is known to convert DNA breaks into ssDNA gaps during the repair of AID-generated dU:dG mismatches. Mice and humans with inactivating mutations in *EXO1* have significant impairments in CSR (Nicolas et al., 2018). An ES threshold of one standard deviation from the mean was chosen to select genes for this analysis and may explain the low number of pathways significant after multiple-testing correction.

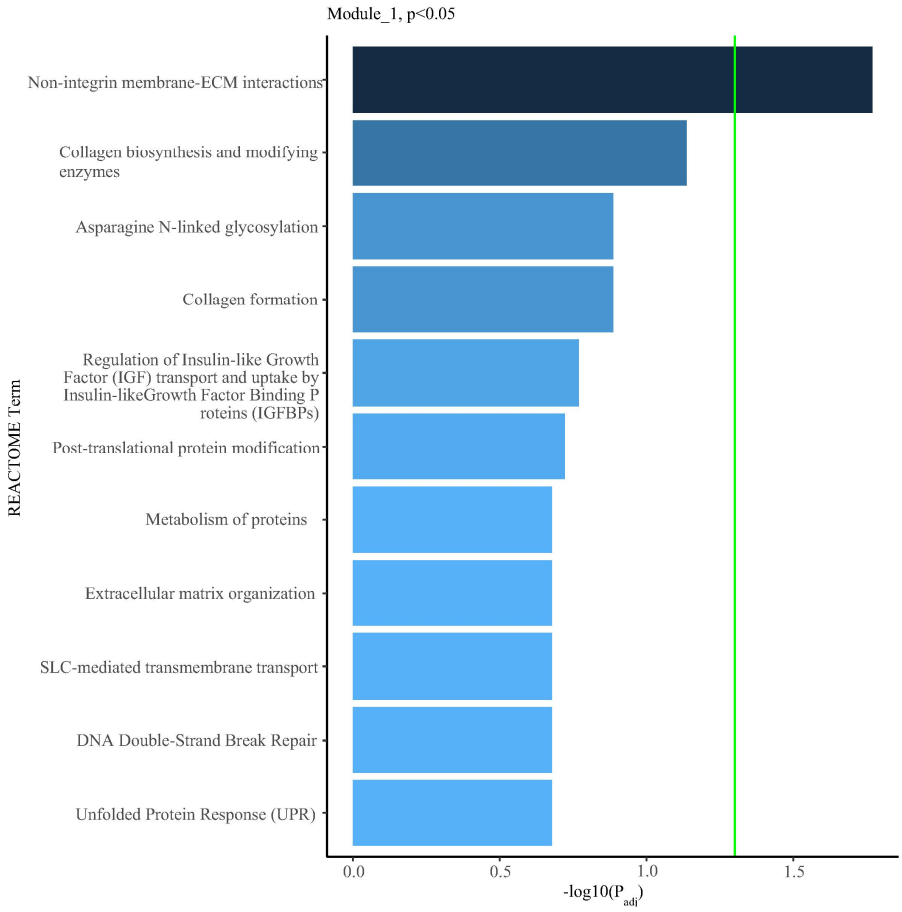


Figure 61. Module 1 associated pathways.

Non-AIR genes significantly associated with module 1 score and with large effect size greater than the mean ± 1 standard deviation, were input into the XGR tool to identify enriched REACTOME terms. The threshold of p_{adj}=0.05 is shown in green.

4.3.3.2 Module 3: Identification of a Putative IgD$^+$ Plasmablast Population Associated with Mortality in Sepsis

Module 3 score was significantly elevated in sepsis compared to health and within sepsis, module 3 score was significantly elevated in 7-day non-survivors at admission to ICU [**Figure 45**, **Figure 48**]. In health and 7-day non-survivors, module score was stable. However, in sepsis survivors and LD non-survivors module 3 score increased significantly over disease course [**Figure 47**]. Module 3 PC1 loading scores are similar across features, thus, all constituent features are well represented by the eigenvector. These include measures of SHM in IgM$^+$/IgD$^+$ BCRs and the proportion of the repertoire comprised of

IgA1 and class-switched isotypes. Elevated IgM^+/IgD^+ SHM was detected across all IGH VDJ regions (CDR3, framework region *etc.*). Non-silent mutations were increased compared to silent mutations indicative of antigen driven clonal affinity selection in response to infection [**Figure 62**]. Plasmablast fraction and module 3 score were positively correlated, therefore it is likely module 3 is a plasmablast driven response in sepsis [**Figure 54**, **Figure C. 3**].

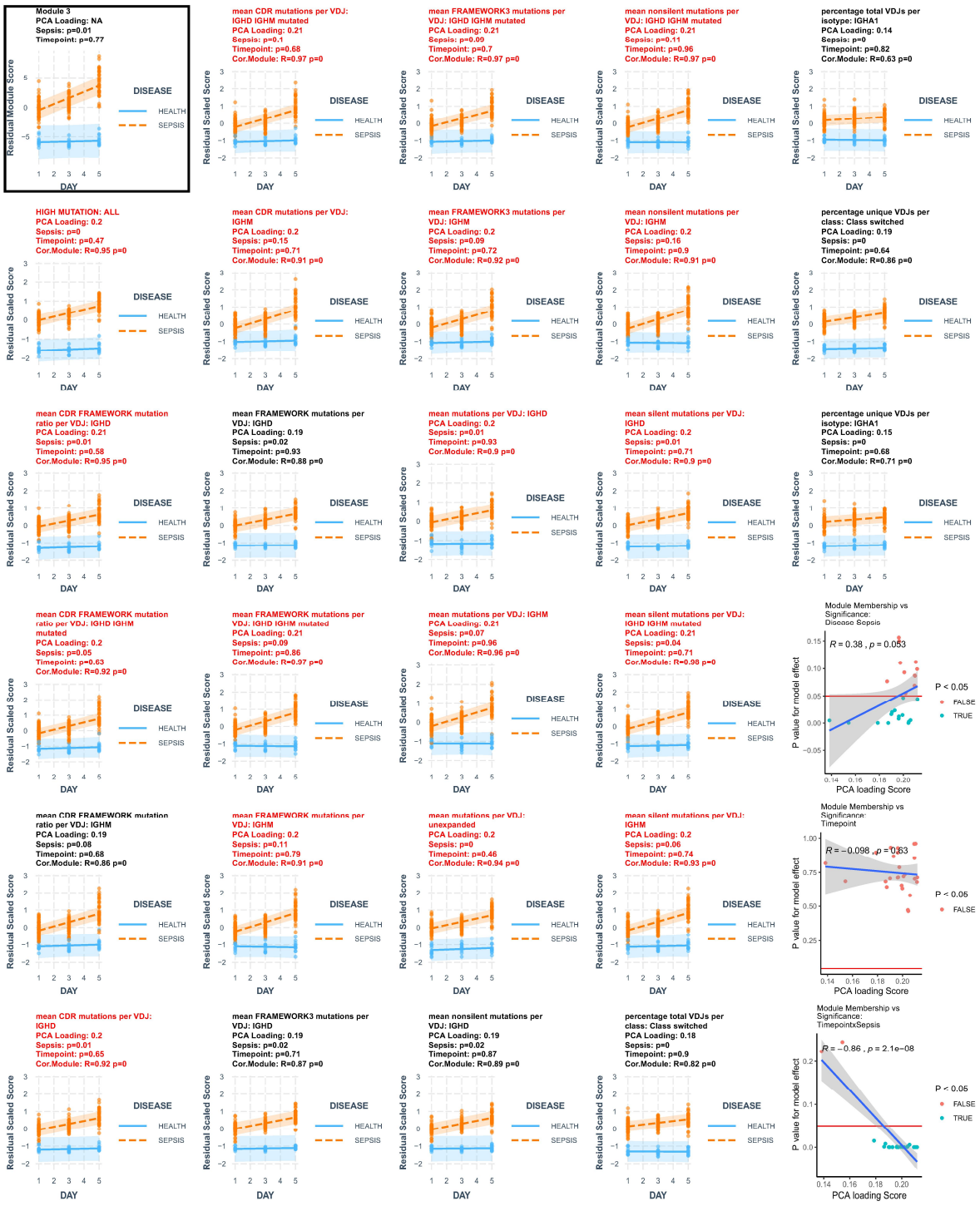


Figure 62. Module 3 is elevated in sepsis and increases significantly during disease course.

Top left shows the partial residual plot for BCR module 3 using the D*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 3 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). 95% confidence intervals are shown.

Mature antigen naïve B cells are typically IgM^+IgD^+ and BCR sequencing data has shown that in health, IgM^+ and IgM^+IgD^+ peripheral B cells typically show little-to-no SHM, which does not significantly differ between isotypes (Kitaura et al., 2017, Klein et al., 1998a). This aligns with the healthy controls included in our study [**Figure 62**, **Figure 63**.**A-F**]. However in sepsis, the level of $IGHM^+$/$IGHD^+$ SHM was significantly elevated compared to health as captured by module 3 score, and SHM was significantly higher in $IGHD^+$ BCRs compared to $IGHM^+$ BCRs. Notably, for both isotypes, SHM frequencies approached that of class-switched isotypes at the later time points [**Figure 62**, **Figure 63**.**C**]. By comparison, SHM in $IGHA1^+$/$IGHA2^+$ BCRs was not significantly elevated in sepsis and for the remaining class-switched isotypes, a small significant increase in SHM was only observed at the later time points.

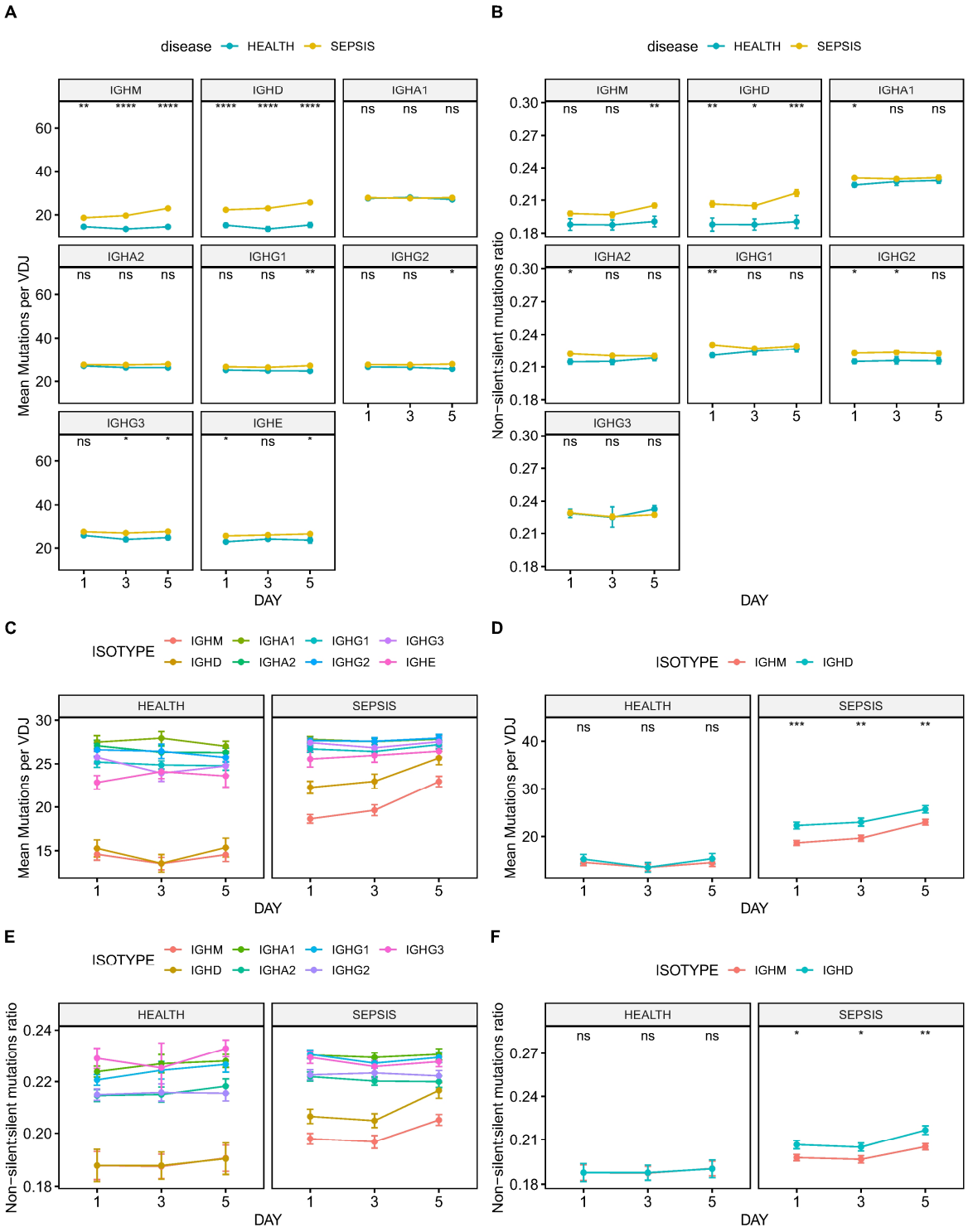


Figure 63. SHM in non-class switched and class-switched isotypes in sepsis.

A and **C)** Mean total number of mutations for each immunoglobulin isotype grouped by disease status. **B** and **E)** Ratio between non-silent and silent mutations for each immunoglobulin isotype grouped by disease status**. D** and **F)** Significant increase in SHM and the ratio of non-silent to silent mutations in IgD$^+$ BCRs compared to IgM$^+$ BCRs. For each plot the mean and 95% confidence intervals are shown.

Research has shown that a small proportion of mucosal B cells from nasopharyngeal lymphoid tissues express only IgD; the result of unconventional CSR. These IgD-secreting plasma cells may enter general circulation to colonise distal mucosal districts. IgD-only B cell populations bind commensals and pathogens inhabiting human upper respiratory mucosae including *Haemophilus influenza* and *Moraxella catarrhalis*. They expand in response to respiratory infections by pathogens such as *Mycobacterium tuberculosis* and in diseases of chronic lung inflammation like asthma. They may also be stimulated by superantigens (Gutzeit et al., 2018, Seifert et al., 2009). IgD^+ antibodies have several unique properties; they are highly reactive to self-antigen, in some cases more than 50% of antibodies from IgD-only B cells bind autoantigens. They also show extensive SHM and may accumulate more mutations than any other B cell population. This has led to the hypothesis that they are sequestered within germinal centre reactions (Koelsch et al., 2007, Arpin et al., 1998, Liu et al., 1996).

4.3.3.2.1 IgD and IgM SHM Distribution

To understand module 3 biology and explain the distribution of SHM observed in the CAP sepsis cohort, I hypothesised the existence of a dynamic IgD^+ plasma cell response in sepsis, initiated in the respiratory lymphoid tissue and detected in peripheral blood due to leaky endothelial barriers, allowing B cells to disseminate more easily into the circulation. To investigate this disease model, the SHM distribution for all IgM^+/IgD^+ VDJ sequences was visualised [**Figure 64**]. Density plots show a clear bimodal distribution in SHM for both isotypes, indicating the existence of at least two distinct populations; a low-mutation naïve B cell population and high-mutation plasmablast or non-class switched memory cell population. Therefore, BCRs were classified based on their mutational load (greater or less

than 17 mutations) to characterise populations separately. No BCRs common to both isotypes were detected, likely due to insufficient read depth.

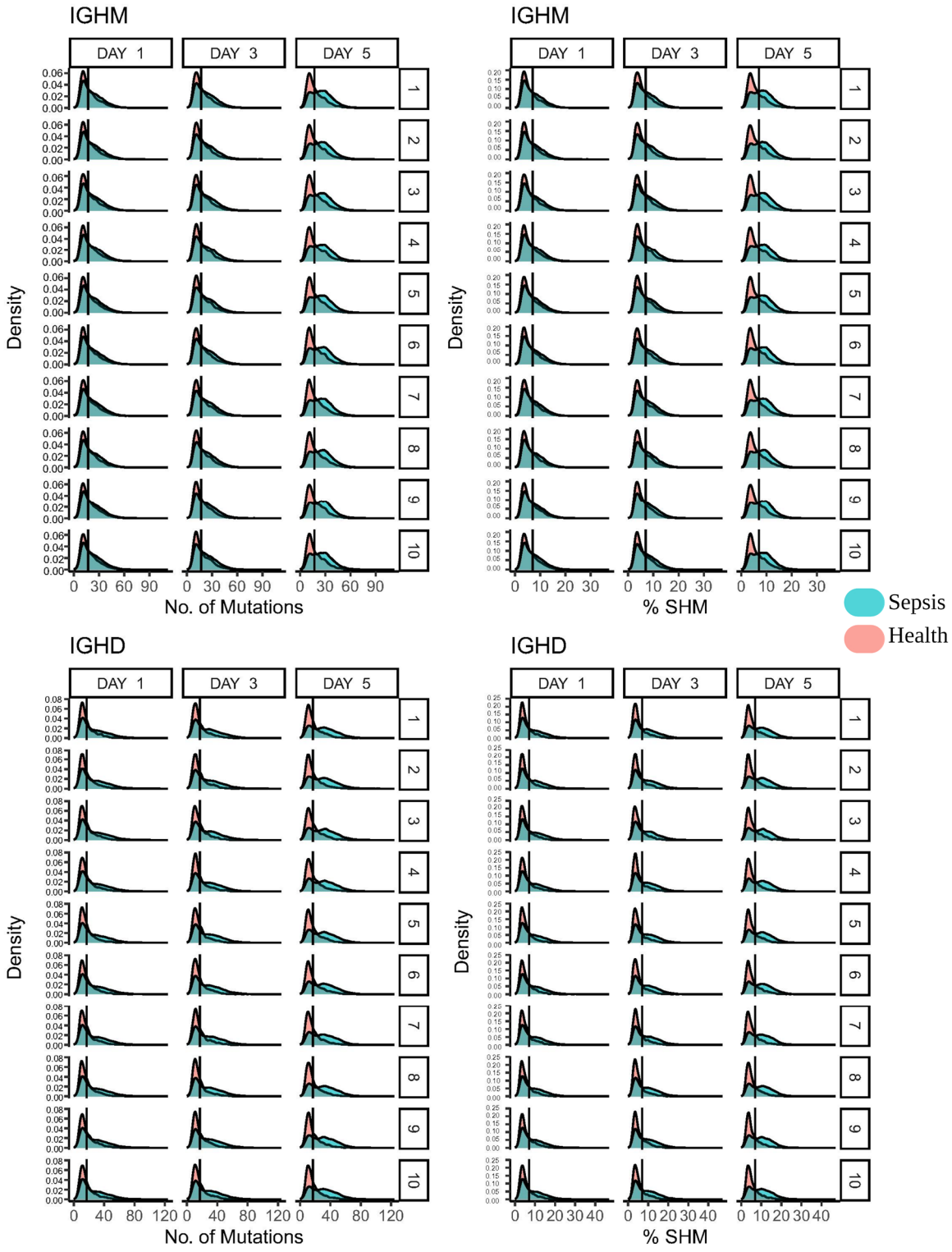


Figure 64. The distribution of SHM for IGHM and IGHD BCR sequences reveals two peaks in mutational load.

The distribution of BCR SHM is plotted using both the total number of mutations and percentage mutation (normalised for read length). Each individual repertoire was subsampled to account for differences in read depth and the distribution was plotted across all individuals. The distributions for 10 replicates are shown and are grouped by disease status and time point. The two peaks are indicative of distinct cellular populations, naïve $IgM^{+}IgD^{+}$ B cells, and $IgM^{-}IgD^{+}$ and $IgM^{+}IgD^{-}$ plasmablast and memory cell populations. A threshold of 17 mutations and 7% SHM is marked.

As expected, the mean level of SHM in the 'antigen naïve' low-SHM population was almost identical between disease and health (<1 mutation difference) [**Figure 65.A**]. The reported statistical significance may result from imprecise classification of sequences. There was no statistically significant difference in mean SHM between IgM and IgD in this group suggesting the BCRs have a common origin, dual expression on antigen naïve B cells [**Figure 65.B**]. The naïve population was depleted in sepsis, consistent with the overall B cell depletion shown in the CIBERSORTx analysis [**Figure 65.C-D**, **Figure 52**]. After normalising SHM to correct for BCR VDJ length, the percentage of SHM per BCR in the naïve population was consistent with the estimate of <5% reported in Kitaura et al., 2016 [**Figure 64**].

In the high-SHM population, both IGHM and IGHD showed significantly greater SHM in sepsis compared to health, although the increase in IgM was only marginal [**Figure 65.A**]. In sepsis, $IGHD^+$ BCRs had significantly higher SHM than $IGHM^+$ BCRs, evidence of a distinct disease-associated IgD^+ plasma cell population [**Figure 65.B**]. The elevated IgD^+ SHM was significant across all time-points, preceding admission to ICU and coincided with population expansion [**Figure 65.C**]. These findings suggest that the IgD^+ response initiated early in disease, possibly at the site of infection (lungs), and before the development of severe sepsis [**Figure 65.A**, **Table 24**]. $IGHM^+$ BCR SHM was also elevated in sepsis, but to a lesser extent, indicating the expansion of an IgM^+ plasma cell population. In health, there was no significant difference in SHM between the two isotypes, therefore BCRs likely originate from IgM^+IgD^+ memory cells which represent ~15% of the B cell pool, or IgM^+IgD^- which comprise around 5% (Seifert et al., 2015, Klein et al., 1998b).

4.3.3.2.2 IGHV3-30 is Enriched in Highly Mutated IgD$^+$ BCRs

Similar analysis performed by Ohm-Laursen et al. (2018) on sequencing data from asthmatic bronchial biopsies also revealed a bimodal IGHD$^+$ BCR SHM distribution. Some tissue samples had mean IGHD SHM approaching that of IGHG and IGHA akin to **Figure 63.C**. The authors used this as evidence of IgD-only B cell clones in inflamed bronchial mucosa. The SHM in the IgD-only clones was significantly higher than in IgM$^+$IgD$^+$ clones and showed typical patterns of SHM localisation, similar again to the sepsis data.

Both Ohm-Laursen et al. (2018) and Seifert et al. (2009) report enrichment of IGHV3-30 gene usage in IgD-only peripheral B cells and IgD plasma cells. Similarly, we observed a significant enrichment in IGHV3-30 usage in the IgD$^+$ high-SHM (plasmablast) population in sepsis; although not to the extent reported in Seifert et al. (2009) [**Figure 65.I-J**]. The presence of diffuse bilateral lung infiltrates at day 1 was associated with increased module 3 score, further evidence linking module 3 and IgD plasma cell response with immune response to respiratory infections [**Figure 50**].

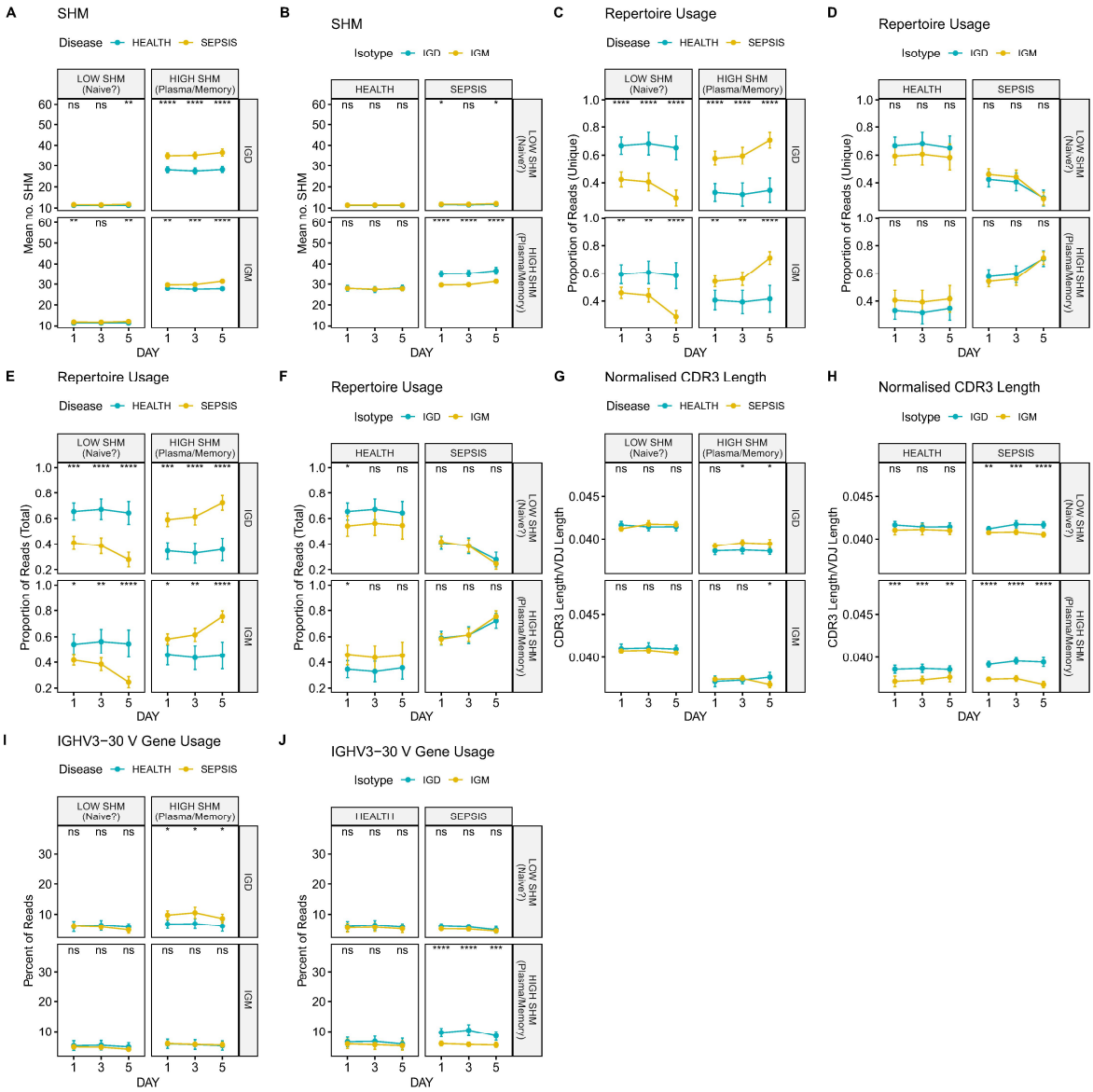


Figure 65. IgM$^+$ and IgD$^+$ BCRs were grouped into naïve and plasma or memory cell populations, according to the number of SHMs.

A, B) Within each group the mean number of SHMs per VDJ was compared across health and disease and between isotypes. **C, D, E, F).** The proportion of reads within each isotype was plotted for the two SHM groups for both total and unique VDJ sequences. **G, H)** The normalised CDR3 length between groups. **I, J)** IGHV3-30 gene usage as a proportion of the total repertoire between groups. The mean and 95% confidence intervals are shown.

Isotype	SHM Group	Disease	Dynamic	Predicted Origin
IgM	Low	Health	Stable	Naïve IgM^+IgD^+
IgM	Low	Sepsis	Decreasing	Naïve IgM^+IgD^+
IgM	High +	Health	Stable	Memory IgM^+
IgM	High +	Sepsis	Increasing	Plasma IgM^+ (Memory IgM^+)
IgD	Low	Health	Stable	Naïve IgM^+IgD^+
IgD	Low	Sepsis	Decreasing	Naïve IgM^+IgD^+
IgD	High +	Health	Stable	Memory $IgM^{+/-}$ IgD^+
IgD	High ++	Sepsis	Increasing	Plasma IgD^+ (Memory $IgM^{+/-}$ IgD^+)

Table 24. Predicted origin of IgM and IgD BCRs in health and sepsis, based on SHM burden and disease dynamics.

4.3.3.2.3 Association with Cell Proliferation

Quantitative trait analysis identified >800 genes associated with module 3 score. One of the most significantly associated genes was syndecan-1 (p_{adj}=0.001, ES=1.36), which, as discussed previously, is a marker for plasma cells and plasmablasts and is required for the survival of antibody-secreting cells (McCarron et al., 2017). Published human data has shown that only a low percentage of IgD^+IgM^- B cells express SDC1 (Rouaud et al., 2014). However, the observed association with module 3, likely reflects the contribution of both IgM and IgA1 B cells to module 3 score. MZB1, another plasma cell marker, was also significantly associated with module 3 score (p_{adj}=1.61e^{-24}, ES=1.7).

Pathway analysis identified multiple enriched pathways centred on the cell cycle, DNA replication and DNA repair [**Figure 66**]. This agrees with literature; pathways that distinguish plasmablasts from other B cell developmental stages include cell proliferation and adhesion (Grasseau et al., 2021, Klein et al., 2003).

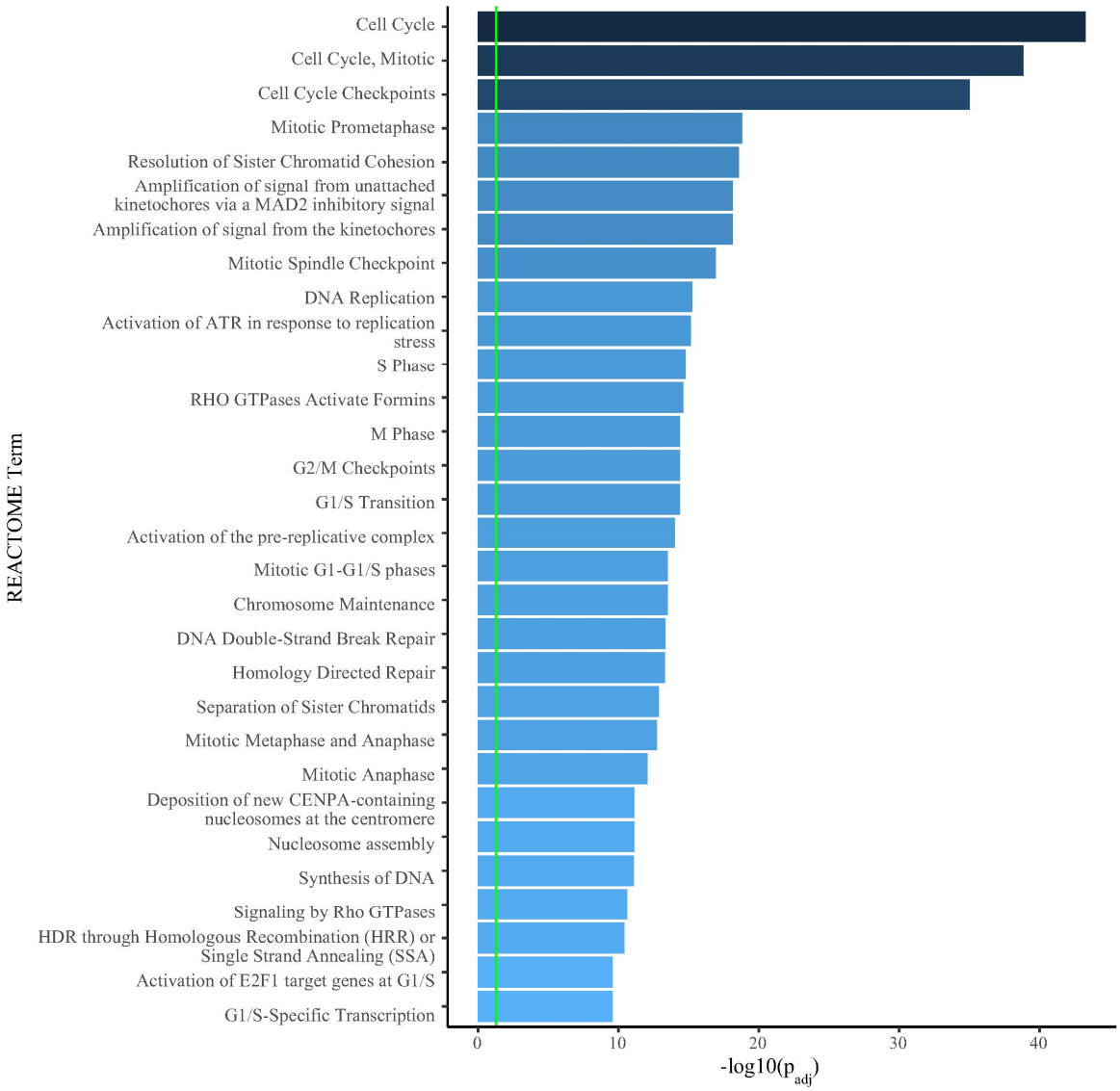


Figure 66. Module 3 associated pathways.

Non-AIR genes significantly associated with module 3 score were input into the XGR tool to identify enriched REACTOME terms. The threshold of p_{adj}=0.05 is shown in green.

4.3.3.3 Clonal Expansion and Clonal Diversification in Sepsis

4.3.3.3.1 Module 13: Increased Intra-clonal Diversification in Sepsis

Module 13 represents BCR intra clonal-diversification due to SHM in IgD, IgM, IgA1, IgA2, IgG1, and IgG2 isotypes and IgM and IgD B cell clonal expansion. Both metrics are derived using Rényi Indexes, which are inverse measures. Thus, the significant decrease in module 13 score in sepsis is associated with increased clonal expansion and clonal diversification.

No association with sepsis mortality was observed and module score did not change significantly across time. IgG1 and IgG2 clonal diversification in conjunction with increased IgG1 usage (**4.3.3.1**) may be particularly beneficial in sepsis as IgG2 is the dominating subclass of antibodies specific to pneumococcal C–polysaccharide (Hjelholt et al., 2013). Module 13 was associated with a significant decrease in plasma cell fraction; this can be interpreted as a positive correlation between plasmablast fraction and intra-clonal diversification [**Figure C. 3**]. Naïve B cells typically undergo little-to-no expansion, therefore elevated clonal expansion in the IgM/D compartment in sepsis is further evidence of IgM^+ and IgD^+ plasmablast populations in sepsis (DeWitt et al., 2013).

Quantitative trait analysis identified only 25 genes significantly associated with module score, 9/25 were AIR genes. Of these, all IGHV genes are part of the IGHV4 family: IGHV4-34 (p_{adj}=0.0041, ES=-0.71), IGHV4-28 (p_{adj}=0.00083, ES=-1.33), IGHV4-55 (p_{adj}=0.00030, ES=-1.40) and IGHV4-31 (p_{adj}=0.013, ES=-0.94). IGHV4-34 germline encoded antibodies are intrinsically autoreactive, recognise carbohydrates expressed on erythrocytes and preferentially bind antigens expressed by commensal bacteria. BCRs expressing IGHV4-34 expand after infection with lymphotropic viruses including CMV and EBV, both of which reactivate relatively frequently in sepsis (Schickel et al., 2017, Mockridge et al., 2004, Zelazowska et al., 2020, Sutton et al., 2009).

AIR gene expression quantified using bulk RNA sequencing data can be inaccurate due to the challenges of mapping short incomplete contigs to the highly repetitive BCR IgH loci using conventional RNA-seq pipelines. However, there was a significant correlation between IGHV gene usage comparing targeted repertoire amplification and bulk RNA sequencing gene expression for both IGHV4-34 and IGHV4-31 ($p<0.05$) although strength of correlation was weak ($R<0.35$). There was no significant correlation for the remaining module-13

associated IGHV genes (IGHV4-28 and IGHV4-55) [**Figure 67**]. Both IGHV4-28 and IGHV4-55 comprise a very small proportion of the total repertoire so the lack of correlation may be attributed to low read depth and stochastic sampling. Despite the lack of correlation between bulk and targeted IGHV4-55 usage, module 13 was negatively correlated with module 5 which captured IGHV4-55 gene usage [**Figure 49**]. An association between IGHV4 gene usage and clonal expansion and diversification across individuals may suggest a convergent response, perhaps to common reactivating viruses.

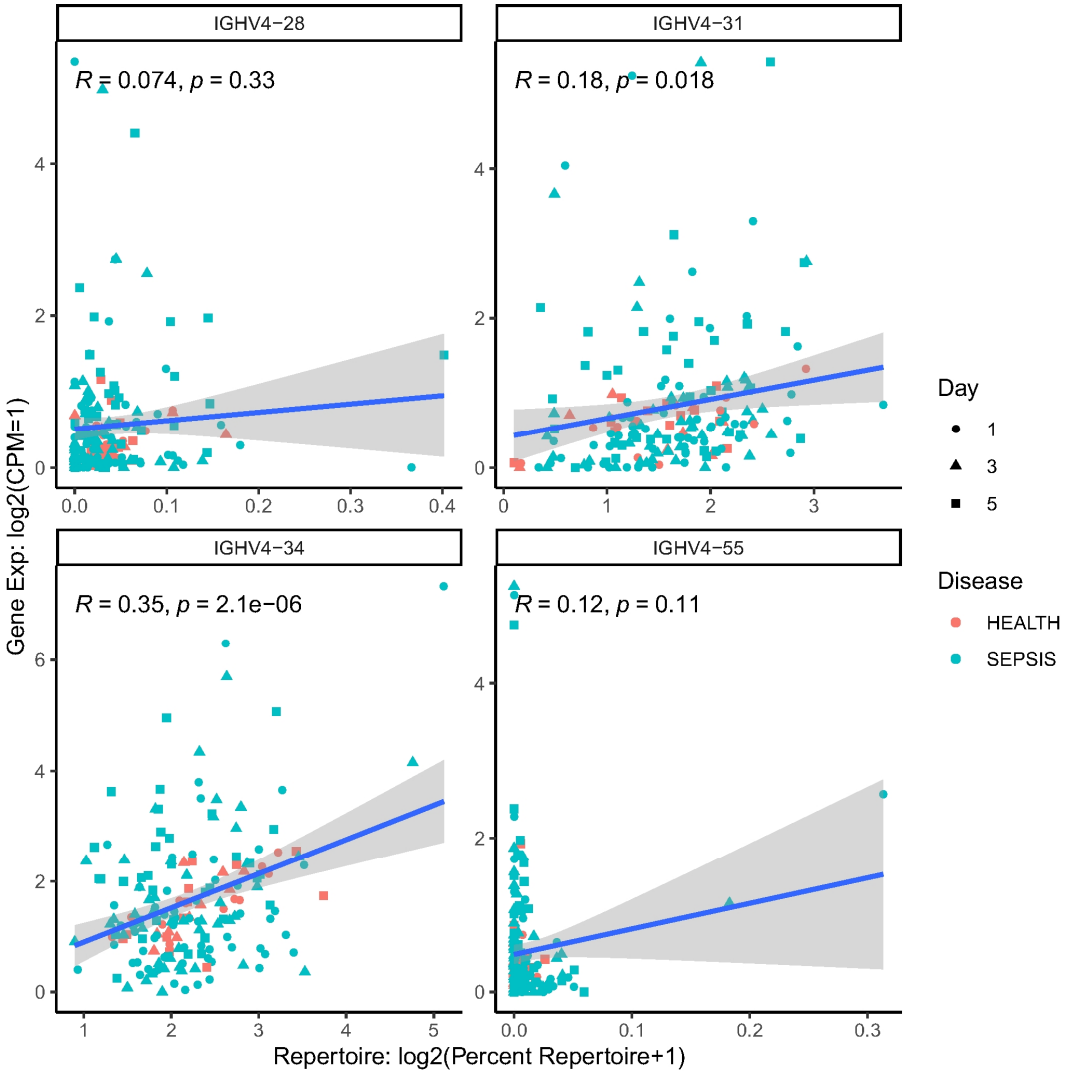


Figure 67. Comparing IGHV gene usage determined using either bulk RNA sequencing gene expression or targeted multiplex PCR amplification of the repertoire.

R and p values are reported for the Pearson correlation coefficient.

RASSF6 which is a putative tumour suppressor gene and member of Ras association domain family genes was significantly negatively associated with module score (p_{adj}=0.0043, ES=-2.366). RASSF6 is frequently suppressed in human cancers, plays a role in malignant progression of solid tumours and is involved in regulating the cell cycle, inhibiting tumour growth and inducing apoptosis. In lymphoblastic leukaemia, *RASSF6* undergoes epigenetic silencing and methylation is associated with worse outcome (Younesian et al., 2017, Hossain et al., 2018, Hesson et al., 2009). Under infection, where B cell proliferation, diversification and expansion are promoted, upregulation of the RASSF6 tumour suppressor gene may help to prevent the development of B cell neoplasms. There were no module 13 associated pathways.

4.3.3.3.2 Module 19: Increased Clonal Expansion in Sepsis

Module 19 also captured measures of clonal expansion; the proportion of the repertoire comprised of the top 5, 10 and 50 clones, the mean and maximum number of unique reads per BCR and mean and maximum total cluster size were all elevated in sepsis. There was no change in module 19 score during sepsis disease course and clonal expansion was not associated with mortality, however module 19 score was significantly positively associated with age. There was a strong negative correlation between module 19 and module 13 score; suggesting modules capture overlapping biology **[Figure 49]**.

Only 88 genes were significantly associated with module 19 score and a single enriched pathway 'Metabolism' was detected using XGR ($p<0.05$) although this was not significant after BH correction for multiple testing. The module 19 associated genes belonging to this pathway are *ALDH1L2* (p_{adj}=0.0016, ES=2.43), *ASS1* (p_{adj}=0.027, ES=3.57), *CAV1* (p_{adj}=0.0034, ES= 1.90), *CD38* (p_{adj}=0.011, ES=1.28), *CHAC1* (p_{adj}=0.0075, ES=2.66),

CHPF (p_{adj}=0.00014, ES=1.44), *GLDC* (p_{adj}=0.0067, ES=1.25), *PYCR1* (p_{adj}= $3.99e^{-5}$, ES= 1.58), *SDC1* (p_{adj}=$3.79e^{-10}$, ES=1.80), and *SLC23A1* (p_{adj}=0.029, ES=3.69), and likely reflect the increased energy consumption of B cells undergoing rapid proliferation and expansion. Interestingly, Chondroitin polymerising factor (*CHPF*) has been shown to be important in proliferation and progression of several different cancers including non-small-cell lung cancer and melanoma (Hou et al., 2019b, Sun et al., 2020).

4.3.3.3.3 Module 16 and 31: Marked Clonal Expansion in IgG in Sepsis

Continuing from **4.3.3.3.1** and **4.3.3.3.2**, significant elevation in module 31 score was observed in sepsis. Module 31 captures specifically measures of IgG1 clonal expansion and diversification, including mean cluster size, the maximum number of unique BCRs per IgG1, IgG1 clonal diversification and the proportion of the repertoire explained by the top 5 and 10 IgG1 clones. Module 16 score was also significantly elevated in sepsis and included additional measures of IgG clonal expansion including the proportion of the repertoire explained by the top 50 IgG1, IgG2 and IgG3 clones. Module 16 score was significantly elevated at day 1 but decreased during disease course. Modules 31 and 16 were the only isotype-specific clonality modules to achieve statistical significance between health and disease suggesting that overall clonal expansion in sepsis is driven in large part, by IgG.

Only 47 genes significantly associated with module 31 were identified; Pleckstrin Homology And RhoGEF Domain Containing G7 (*PLEKHG7*) had the second largest ES and has been previously shown to be upregulated in rare thymus memory IgG^+ B cells (p_{adj}=0.0051, ES=6.40) (Bergkvist et al., 2016) [**Figure 68**]. The long non-coding RNA *CLCA4-AS1* had the third largest ES (p_{adj}=0.00033, ES=6.32). *CLCA4-AS1* has been linked to serum IgG

glycosylation using GWAS, and influences IgG effector function by modulating Fc receptor binding (Lauc et al., 2013). *LILRB1* expression had a strong negative effect on module 31 score (p_{adj}=0.0039, ES=-1.45). *LILRB1* is a member of the Leukocyte immunoglobulin (Ig)-like receptors and is a receptor for a broad range of classical and non-classical HLA class I molecules and uncharacterised molecules produced by *Staphylococcus aureus,* Escherichia coli and human cytomegalovirus which suppress immune cell activity. LILRB1 engagement on B cells impairs B cell function including proliferation, reducing antibody production (Abdallah et al., 2021, De Louche and Roghanian, 2022).

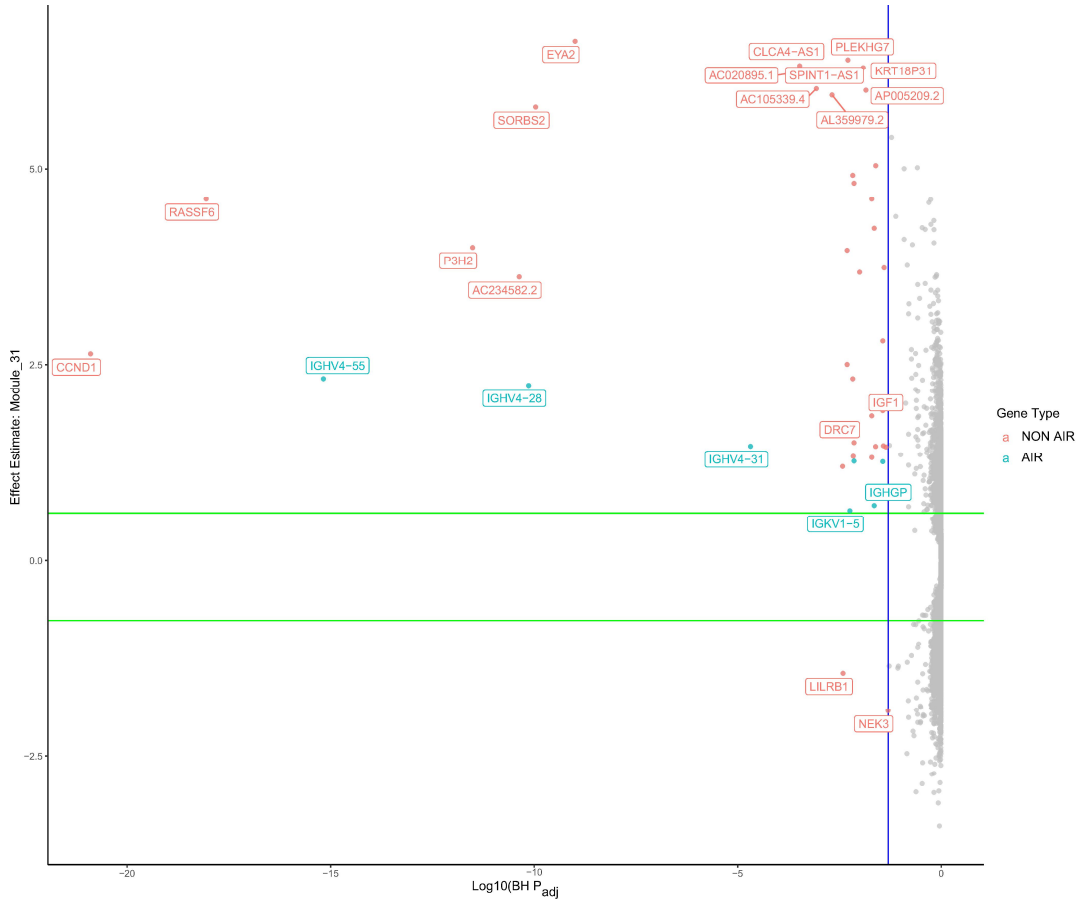


Figure 68. Quantitative trait analysis identified 47 genes significantly associated with BCR module 31 score.

Linear mixed effects models were used to identify genes significantly associated with module 31 score, using bulk RNA sequencing gene expression data. P values were corrected for multiple testing using the Benjamini-Hochberg method. The log10(p=0.05) is shown by the blue line and the mean effect size ± 1 standard deviation are shown in green. Genes are annotated as either adaptive immune repertoire genes (AIR) or non-AIR genes.

4.3.3.4 Module 29: Poor IgM Antibody Response is Associated with Mortality in Sepsis

Module 29 captures IgM and IgD isotype usage, including the percentage of total or unique VDJs for unmutated IgM/IgD BCRs, mutated IgM/IgD BCRs and overall IgM BCRs. It also includes the proportion of the repertoire comprised of low SHM BCRs indicating that these BCRs are predominantly IgM/D and of naïve B cell origin. Module 29 score was significantly decreased in sepsis and decreased further during disease course while remaining stable in health. Furthermore, at day 1, module 29 score was significantly decreased in the 7-day non-survivors relative to survivors suggesting it may be a marker of poor outcome in sepsis. Although not significant, late-death non-survivors also had a lower mean module 29 score compared to sepsis survivors.

Analysis of matched bulk RNA sequencing immunoglobulin isotype gene expression, revealed that absolute *IGHM* and *IGHD* expression while lower in sepsis than in health, increased during disease course, but not to the extent of the class-switched isotypes [**Figure 55**]. Therefore, a better interpretation of module 29 is reduced IGM/D usage *relative* to class-switched isotypes. The association with mortality on day 1 suggests maintenance of the IgM antibody response in sepsis, rather than biased expansion of class-switched isotypes is important for good prognosis. These findings are in agreement with existing literature; Bermejo-Martin et al. (2014), Martinez et al. (2021b) report low serum IgM is a risk factor for mortality in sepsis and treatment with IgM-enriched IgG (IgGM) helps to mitigate that risk. This has also been shown in paediatric sepsis (Haque et al., 1988). A drop in IgM is characteristic of patients who transition from severe sepsis to septic shock (Giamarellos-Bourboulis et al., 2013). Interestingly, low IGM is a risk factor for the development of sepsis after rituximab treatment but not for other infections (Ezhuthachan et al., 2021).

4.3.3.5 Module 26: Biased IGHV5 Gene Usage is Associated with Mortality in Sepsis

Module 26 was the only BCR module significantly elevated in sepsis 7 day non-survivors at all 3 time points and increased significantly during disease course [**Figure 69**]. In both health, sepsis survivors and late-death non-survivors module 26 score was stable. Module 26 represents IGHV5 gene usage (IGHV5-51 and its paralog, IGHV5-10-1) and the overall usage of the IGHV5 gene family across all immunoglobulin isotypes. The only other IGHV5 family member is IGHV5-78. Bulk RNA-seq IGHV5-10-1 expression was not significantly associated with module score after BH correction (p=7.7e-05, p_{adj} =0.76, ES= 0.81), nor was IGHV5-51. However, as these genes are paralogs this could result from multi-mapping or mismatching during pre-processing.

Biased IGHV5-51 and IGHV5-10-1 usage has been linked to the autoimmune diseases multiple sclerosis and celiac disease (Snir et al., 2015, Lomakin et al., 2014, Iversen et al., 2019). Quantitative trait analysis identified only one gene associated with module 26 score, AL671883.3 (p_{adj}=0.03, ES=2.3). AL671883.3 (Ensemble ID: ENSG00000285647) is a novel transcript which localises to the HLA-B region (MHC class I) on Chromosome 6: 31,367,057-31,376,088 and has been linked to ankylosing spondylitis and psoriasis (Ko et al., 2022, Stuart et al., 2022).

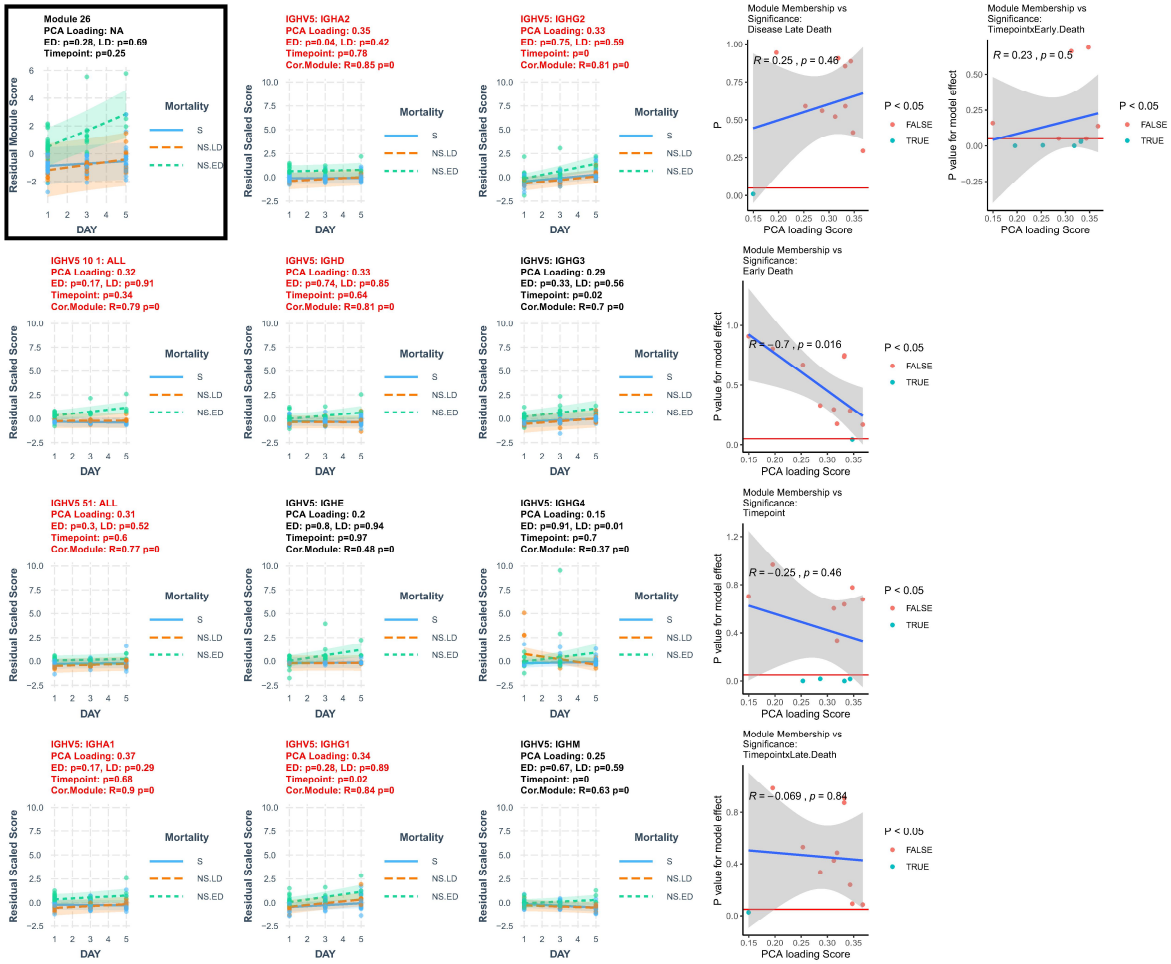


Figure 69. Partial residual plots depicting biased IGHV5 gene usage associated with sepsis mortality, controlling for covariates.

Top left shows the partial residual plot for BCR module 26 using the M*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 26 cluster, modelled using the same M*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are five scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (mortality group, time point and disease time point interaction). Results are reported to two decimal places.

4.3.3.6 Sample Stratification

To identify subgroups of patients with similar repertoires, hierarchical clustering was performed using the imputed BCR repertoire feature matrix. Firstly, a distance matrix was calculated using the Euclidian distance, then, agglomerative hierarchical clustering using Ward's linkage method was performed. Dynamic Hybrid Tree Cut was used to determine

the optimal number of patient clusters with a minimum cluster size of 20 [**Figure 70**]. The dendrogram reveals two distinct branches, with one containing only sepsis samples, while the other contains a mix of patient and control samples. There are clear instances where longitudinal samples from the same patient are highly similar, demonstrating the personalised nature of the BCR repertoire. Such reproducible clustering is likely driven by the inclusion of IGHV gene usage in the feature matrix.

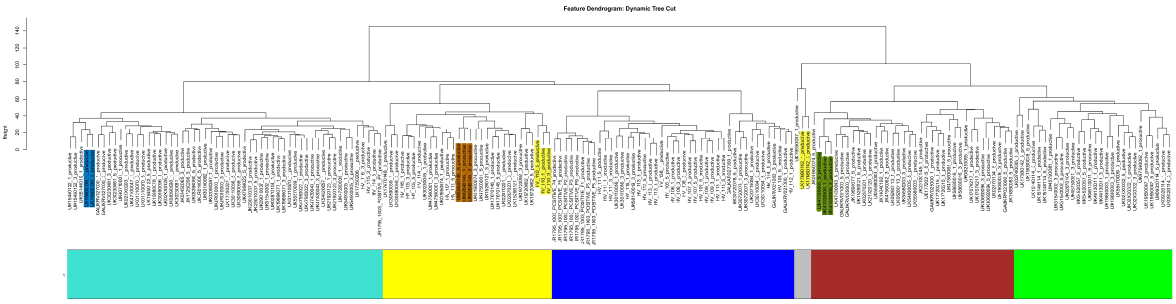


Figure 70. Patient stratification by hierarchical clustering of the imputed BCR repertoire feature matrix.

Patients were clustered using the imputed BCR repertoire feature matrix. Examples of good clustering of longitudinal patient samples have been highlighted, demonstrating the personalised nature of the adaptive immune repertoire.

PCA was also performed on the BCR repertoire feature matrix and PC1 and PC2 reveal clear separation of healthy volunteer samples from sepsis patients consistent with hierarchical clustering. However, PC1 and PC2 cumulatively explain only ~25% of the observed variance [**Figure 72.A**]. Within sepsis patients, there was no clear stratification based on sepsis mortality using PC1 and PC2. To identify BCR repertoire features driving separation between health and sepsis in PC1 and PC2, feature loading scores were visualised [**Figure 71**]. Reassuringly, the features with large loading scores in both PC1 and PC2 included SHM (**4.3.3.2, 4.3.3.3**), clonal expansion particularly for IgG (**4.3.3.3.3**), IGHM IGHV4 family gene usage and interestingly the proportion of B cells in the top 5 clones

which have evidence of V gene replacement (*e.g.* they share a stem sequence but have different V genes).

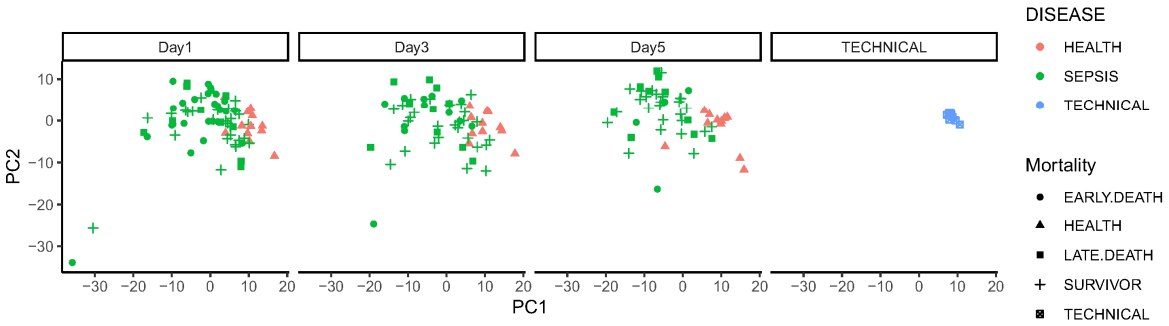


Figure 71. BCR Principal Component Analysis.

PCA was applied to the scaled imputed BCR repertoire feature matrix directly. The first two components are plotted and points are coloured according to disease status. The shape corresponds to the sepsis mortality group. Clear separation of healthy volunteer and sepsis samples is observed.

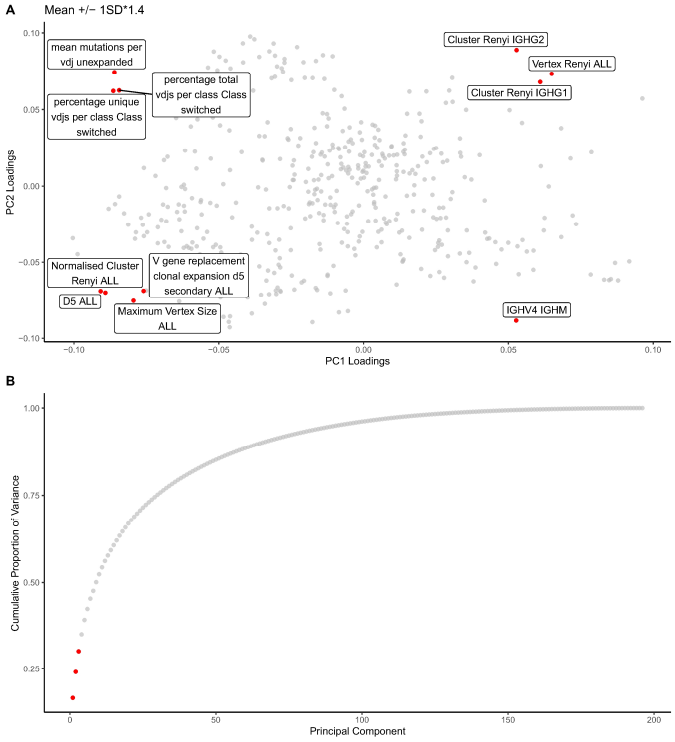


Figure 72. Separation between health and disease is driven by a subset of BCR repertoire features.

PCA was applied to the scaled imputed BCR repertoire feature matrix directly. **A)** PC1 and PC2 feature loading scores were plotted to identify the subset of features which drive separation of health and sepsis in **Figure 71**. Features are highlighted if their loading score is greater or less than the mean ± 1 SDx1.4 for both components. **B)** Cumulative proportion of variance explained by each PC. Points are highlighted if each PC explains more than 5% of variance in the data.

4.3.4 Longitudinal B cell Clonal Dynamics

4.3.4.1 Introduction to Network Structure and Centrality

To understand B clonal evolution and dynamics in sepsis, I applied the computational pipeline MRD Assessment and Retrieval Code in pYthon (MRDARCY) (Bashford-Rogers et al., 2016). In brief, productive BCR sequences from all longitudinal samples per patient were clustered using CD-HIT allowing for 8 base-pair mismatches in sequence identity due to SHM. BCR clusters are termed clones and represent B cell lineages. Of the study participants, 60 individuals including 12 healthy volunteers had BCR repertoire data for more than one time point and were included in this analysis.

BCR networks were generated separately for each participant and BCR clone containing ≥10 unique VDJ sequences spread across more than more time point (**2.6.4**). Neighbourhood-joining tree estimation was used to construct an undirected phylogenetic tree. For the same BCR clone, an undirected minimum spanning tree (MST) was also constructed and used to calculate different centrality measures, which are one of the fundamental principles of network analysis (Golbeck, 2013) [**Table 25**]. Example networks, which demonstrate how differences in network structure influence centrality measures, are shown in the appendix **Figure C. 4**. Longitudinal BCR networks for two clones from both a sepsis patient and healthy volunteer are shown in **Figure 73**.

Metric	Level	Description
Degree Centrality	BCR (Node-level)	The number of (non-loop) edges connected to each node.
Normalised Degree Centrality	BCR (Node-level)	Same as degree centrality, however the result is divided by $(n-1)$ where n is the number of vertices in the graph.
Eigenvector Centrality	BCR (Node-level)	Measures a node's importance and influence within a network by considering the importance of its neighbour; links from important nodes (measured by degree centrality) are more important than links with unimportant nodes (low degree centrality). Ranges between 0-1 with 1 as the most influential node.
Betweenness Centrality	BCR (Node-level)	Measures how important a node is to the shortest paths through a network. Node centrality is defined by the number of shortest paths going through a vertex.
Centralisation	Clone (Network-level)	Centralisation uses the distribution of a centrality measure to understand the structure of the overall network. If centrality is evenly distributed then the centralisation of a network is low.
Degree Centralisation	Clone (Network-level)	Score is normalised by dividing the graph-level centrality by the theoretical maximum for a graph with the same number of vertices. The most centralised network organisation according to degree centrality is the star-graph. Ranges between 0-1.
Eigenvector Centralisation	Clone (Network-level)	Score is normalised by dividing the graph-level centrality by the theoretical maximum for a graph with the same number of vertices. A high score indicates that a small group of nodes with high importance form a densely connected group.
Betweenness Centralisation	Clone (Network-level)	Score is normalised by dividing the graph-level centrality by the theoretical maximum for a graph with the same number of vertices

Table 25. Summary of node and graph-level network centrality measures.

Definitions sourced from Golbeck (2013) and Csardi and Nepusz (2005).

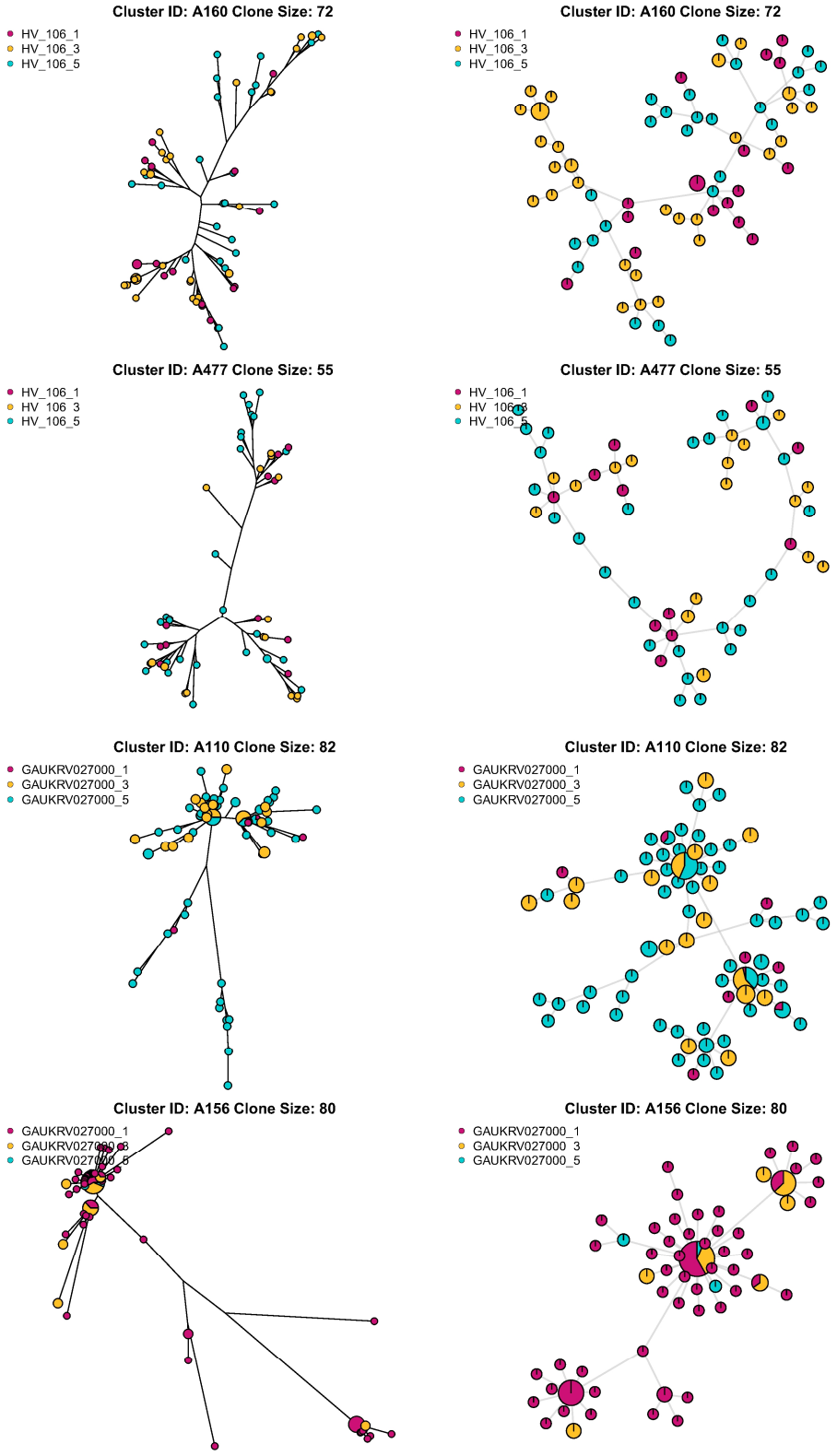


Figure 73. Longitudinal B cell network from a patient with sepsis (LHS) and a healthy volunteer (RHS).

Example phylogenetic trees (left) and minimum-spanning trees (right) for two B cell clones from a sepsis patient (GAUKRV027000) and a healthy volunteer (HV_106). Each node represents a unique BCR sequence; node size is proportional to the number of BCRs and is coloured according to the time point detected.

Degree centralisation was calculated for all clones detected within an individual. Then the mean degree centralisation across clones was compared between health and disease using a Wilcoxon test [**Figure 74.A, B**]. Mean degree centralisation was significantly elevated in sepsis, indicating that clones exhibit a more 'star-like' network structure (as seen in **Figure 73** clone A110 and A156) with B cell clonal evolution focused around a smaller number of central BCR sequences. These central BCR nodes may represent local optima for encountered antigen. The antibody space surrounding the central BCRs is extensively explored via SHM, likely in an attempt to increase antigen affinity, which produces the 'star' structure. There was no significant difference in degree centralisation between sepsis mortality groups. Low degree centralisation in health reflects both the absence of selection pressure exerted by germinal centre reactions and the prevalence of antigen naïve B cells which do not undergo significant expansion. Analysis was replicated using a repeated subsample of clones, to account for differences in read depth. Results were comparable [**Figure 74.E and F**].

The effect of disease status or mortality group and time on the average number of BCR connections (node-level degree centrality) was assessed using a linear mixed effect model with a random effect term for sample ID. There was no significant change in BCR connectedness over time (p=0.95, ES=-0.005), however the average number of connections was significantly higher in sepsis compared to health (p=0.011, ES=0.80) consistent with increased somatic-hyper mutation and clonal diversification [**Figure 75.A**]. Within sepsis those who died early in disease had significantly fewer connections per BCR, compared to survivors (p=0.0055, ES= -0.90) [**Figure 75.B**]. There was no difference between survivors and late-death individuals (p=0.34, E=-0.31). However, this could be a technical artefact,

as the majority of early-death individuals had networks derived from only two time points, therefore the number of available nodes is likely to be fewer.

The distribution of degree centrality was visualised for a random subsample of nodes per individual, for all samples. The majority of BCRs in both health and sepsis have degree centrality ≤3. Therefore changes in the mean degree centrality between health and disease are likely driven by differences in the connectedness of the less frequent central BCRs, rather than an overall increase in connectedness [**Figure 75.C**]. This is consistent with a transition to a more 'star-like' network structure in sepsis where one node is connected to many more BCRs than in health.

Eigenvector centrality is a measure of the influence of a node in a network and assigns relative scores to all nodes based on the principle that connections between high-scoring nodes (with high degree centrality) contribute more to the score of the node in question than an equal number of connections from low scoring nodes. High eigenvector centralisation (network-level measure) indicates that a small group of nodes with high importance form a densely connected group. Eigenvector centralisation was elevated in sepsis compared to health [**Figure 74.D-E**]. A dense connection of high degree centrality nodes suggests that SHM successfully increases BCR antigen affinity in sepsis as secondary BCRs also undergo positive selection, expansion and SHM generating a hub of highly connected 'important' nodes. This structure is visualised in **Figure 73** clone A110 and A156.

Our findings are consistent with B cell network analysis performed by Miho et al. (2019) using CDR3 sequences from sorted murine memory plasma B cells, pre-B cells (pBC) and naïve B cells (nBC). A star-like structure and high centralisation was associated with plasma cell clones (enriched in sepsis). By comparison, a typical tree-like structure was found for

pBC and nBC (enriched in health). Similarly, the authors also show that BCR network measures are robust to extensive subsampling, indicating a widespread shift in clone structure.

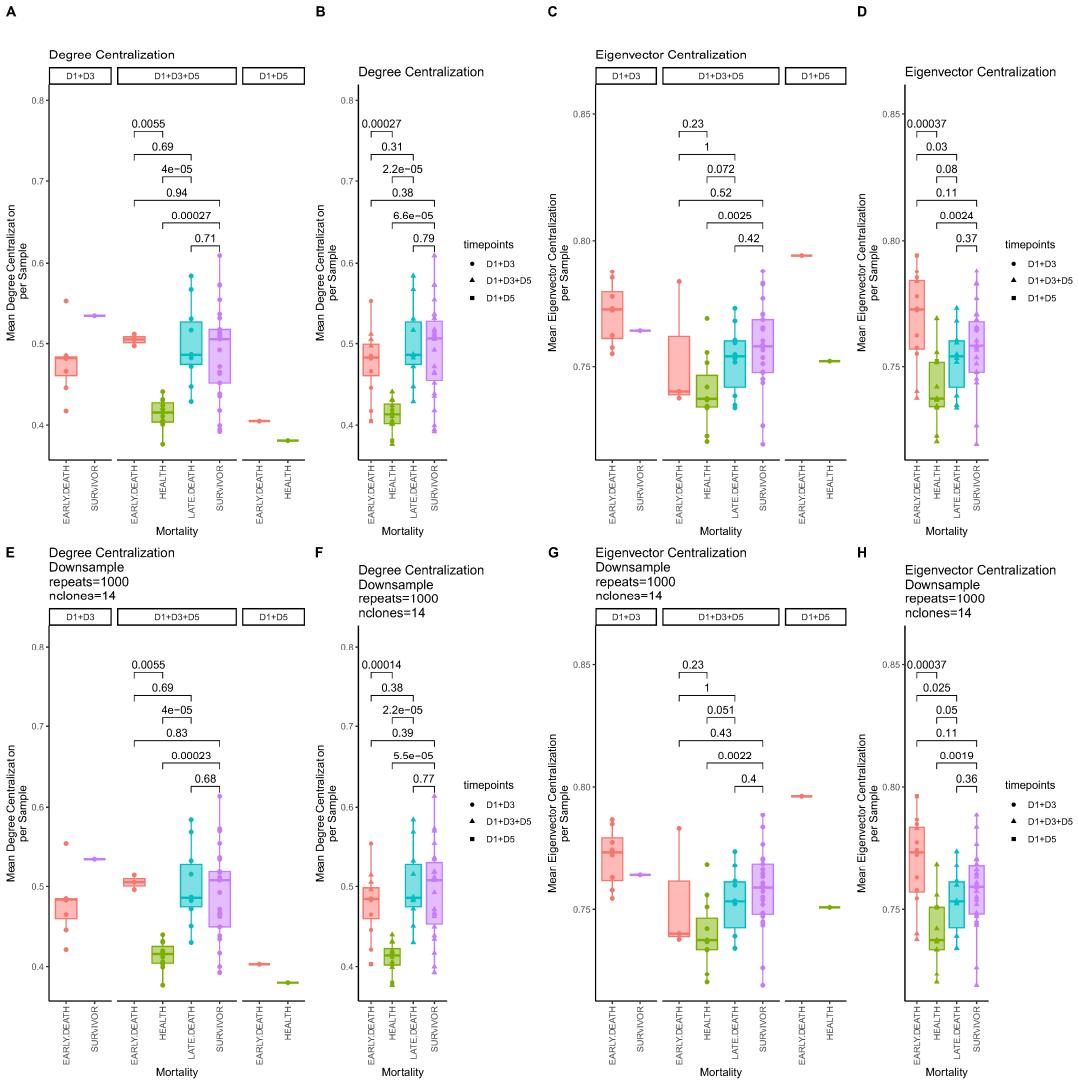


Figure 74. BCR clone network-level centralisation measures grouped by mortality for (A, B, E, F) degree centralisation and (C, D, G, H) eigenvector centralisation.

Upper plots are for the mean of all clones detected per individual while lower plots are for a subsample of 14 clones per individual to account for differences in read depth. **A, C, E** and **G)** Wilcoxon tests are performed separately for clone networks derived from individuals with 2 or 3 available time points while for **B, D, F and H** networks are not grouped by number of time points. Degree and eigenvector centralisation measures are normalised for differences in the number of BCRs, which is expected to be higher in clones from individuals with three rather than two time points. However, the structure of clones derived from adjacent time points is likely to be more connected than those derived using day 1 and day 5 time points only. Due to the small number of samples in each category, it was not possible to evaluate this further.

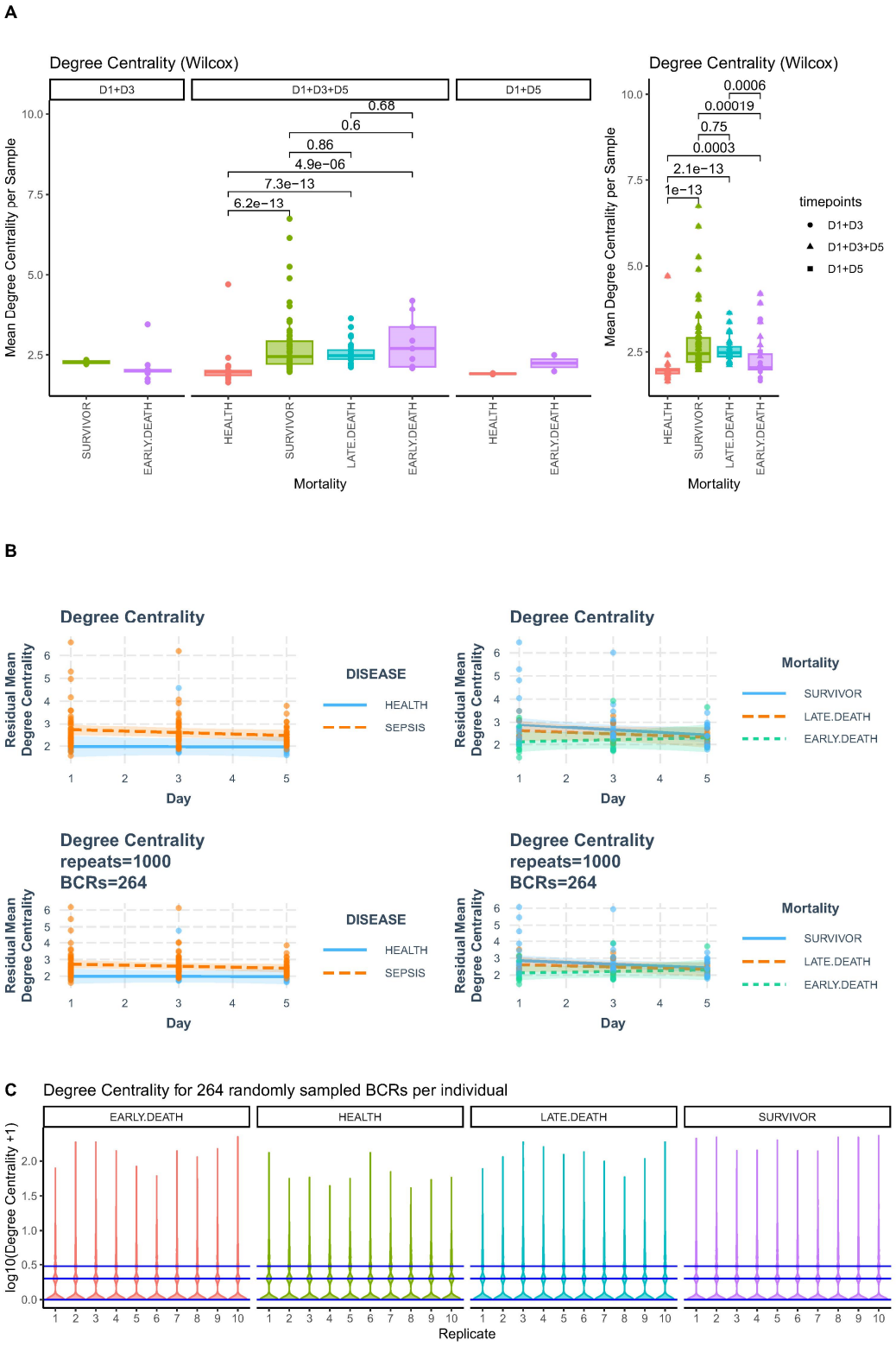


Figure 75. BCR degree centrality.

A) All BCR nodes irrespective of clone, stratified by health and sepsis mortality. P values reported for a Wilcoxon test. **B. Upper)** All nodes grouped by time point detected, for either **Left)** health versus sepsis **Right)** sepsis mortality groups. **B. Lower**) Same as B. Upper but for a subsample of 264 nodes per individual to account for differences in read depth. **C)** Violin plot showing the log(10 Degree Centrality+1) for 10 random subsamples of 264 BCRs from each individual, per mortality group. Thresholds corresponding to degree centrality equal to 1, 2, and 3 are shown in blue.

4.3.4.2 Characterising Central and Persistent BCRs

To characterise persistent BCRs the mean degree centrality was calculated for BCRs detected at one (T1), two (T2) or three (T3) time points. Persistent BCRs are defined as unique VDJ sequences detected at more than one time point. In both health and disease, persistent BCRs have greater degree centrality compared to transient BCRs; the greater the persistence the greater the degree centrality. This suggests that persistent BCRs represent local optima for antigen binding and the surrounding antibody space is extensively explored using SHM [**Figure 76.A-B**]. In both health and disease, persistent BCRs comprise a small proportion of the total clonal BCR population (μ_{T2}=0.047 and μ_{T3}=0.016) [**Figure 76.C**]. Persistent BCRs are significantly enriched in sepsis, consistent with positive selection due to antigen exposure but there was no significant difference in their frequencies between sepsis mortality groups. Degree centrality was significantly higher than health for sepsis persistent clones detected at three time points (5 days) but not those detected at two time points (3 days) (p=0.035 and p=0.5 respectively) [**Figure 76.B**].

Clone-specific central nodes were defined as those nodes (BCRs) with eigenvector centrality equal to one. BCRs with eigenvector centrality not equal to one were termed peripheral nodes. Like persistent nodes, central nodes had greater degree centrality and had more counts than peripheral nodes in both health and sepsis [**Figure 76.D-F**]. Central but not peripheral nodes had significantly greater degree centrality in sepsis compared to health as alluded in **Figure 75.C**. Although persistent and central nodes have similar network properties, the two populations do not fully overlap, only ~50% of clones detected at all three time points were also assigned as central nodes [**Figure 76.F**]. This percentage was markedly lower for clones detected at two time points (~15%). For transient clones ~1.5% were assigned as central nodes.

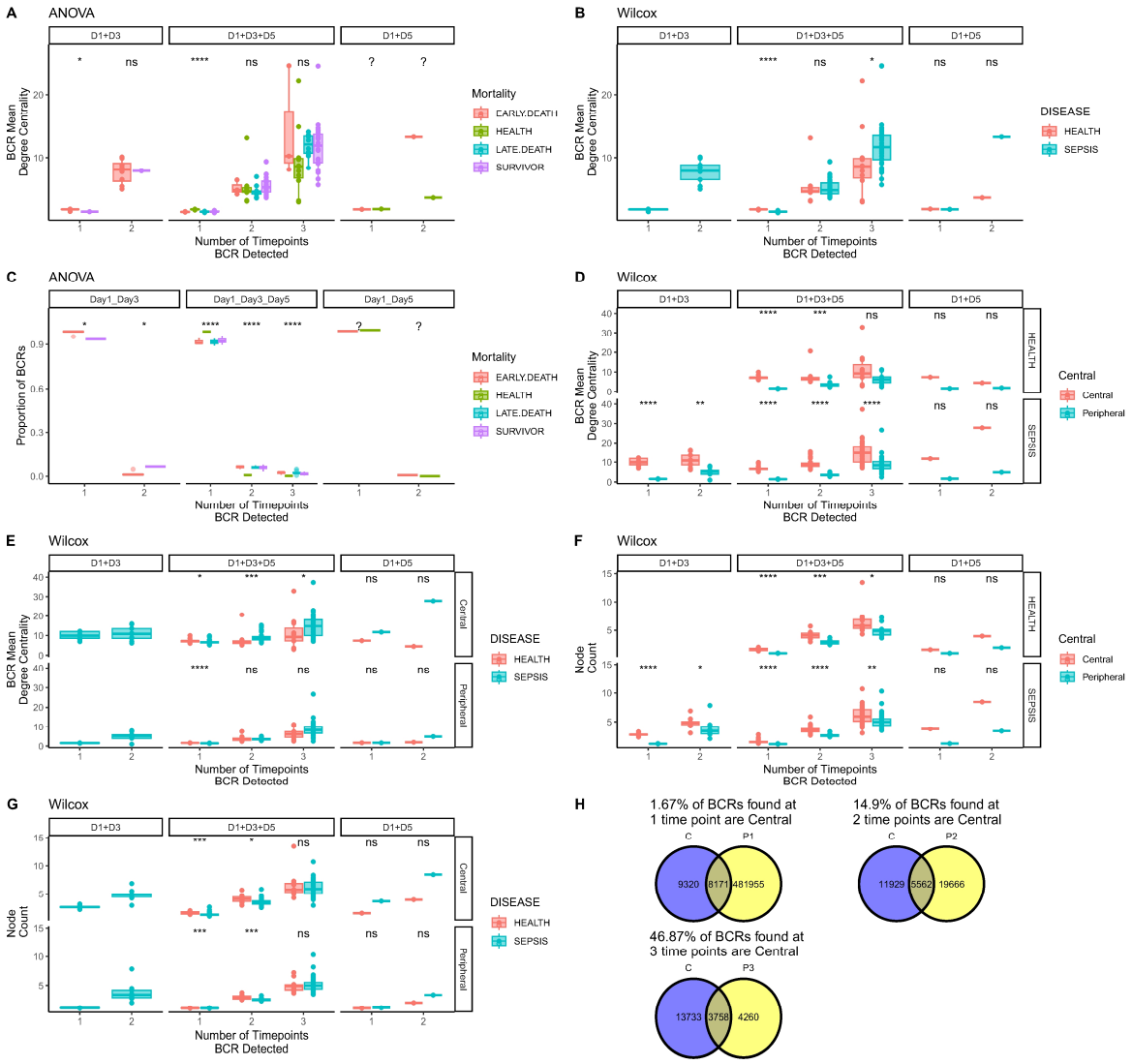


Figure 76. Characterising persistent and central BCRs.

A, B) Mean degree centrality for persistent BCRs. **C)** Proportion off the clonal repertoire comprised of persistent BCRs. **D&E)** Mean degree centrality for 'central BCRs' (determined by Eigenvector centrality). **F&G)** Mean node (BCR) size for central BCRs. **H)** The overlap between persistent and central BCRs.

4.3.4.3 Clonal Overlap

The Jaccard index was used to assess clonal overlap between time points for three transitions: day 1 versus 3, day 3 versus 5 and day 1 versus 5 [**Figure 77**]. All detected clones were used in this analysis, irrespective of clone size. The Jaccard index was significantly elevated in sepsis compared to health for all transitions. However, there was no significant difference between sepsis mortality groups. This is consistent with an

increased percentage of persistent BCR sequences as described in **4.3.4.2**. For both sepsis survivors and late-death non-survivors, adjacent time points had significantly greater clonal overlap than non-adjacent time points indicative of ongoing expansion and selection. This was not the case in health and early-death non-survivors.

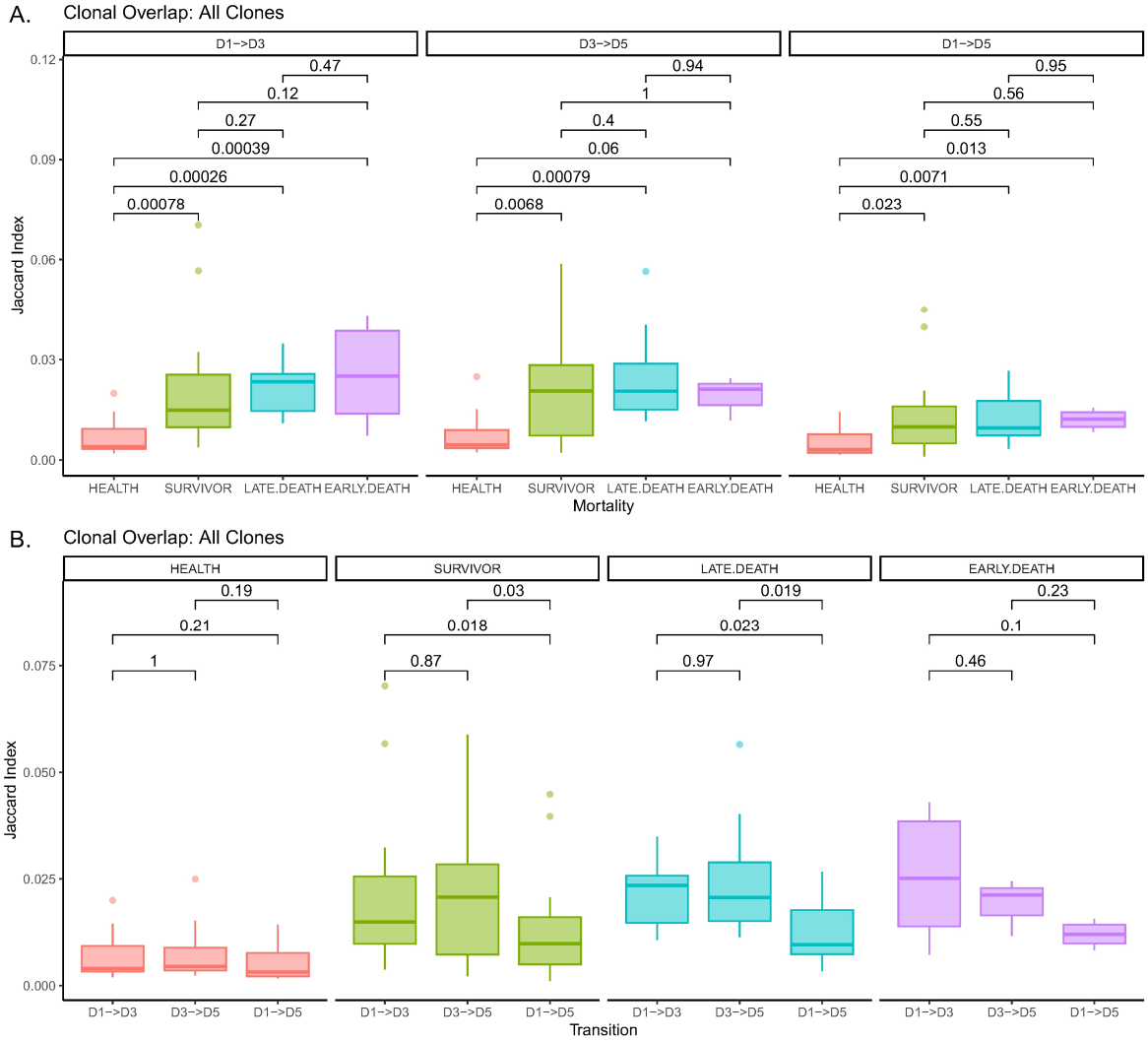


Figure 77. The Jaccard index was used to calculate the clonal overlap between the BCR repertoires at different time points.

To test for a significant difference in mean clonal overlap a Wilcoxon test was used; p values are shown. **A)** Comparing the overlap between different mortality groups for each transition. **B)** Comparing the overlap between each transition across mortality groups.

4.3.4.4 Clonal Dynamics

Clones were classified as increasing, decreasing (dynamic clones) or stable based on the percentage change in repertoire between the first time point and last sequential time point available. All detected clones were used in this analysis, irrespective of clone size. To account for stochastic changes due to resampling I first compared the percentage change in repertoire for all clones in the data set and used this distribution to calculate a mean and standard deviation [**Figure 78.A-B**]**.** Mean change in percentage tended to 0. Clones with percentage change within one standard deviation of the mean were classified as stable clones. The majority of BCR clones were stable across time and there was no significant difference in the number of stable clones between health and sepsis. The proportion of clones which decreased was elevated in sepsis compared to health. Within sepsis, early-death non-survivors had an elevated proportion of decreasing clones [**Figure 78.C**].

The proportion of each clone comprised of every immunoglobulin isotype was calculated and averaged across clones for each individual [**Figure 79.A&B**]**.** In both sepsis and health, stable clones had elevated IGHM and IGHD usage and reduced usage of class-switched isotypes IGHA1 and IGHA2 compared to dynamic clones. The average clone size was also significantly smaller than that of dynamic clones [**Figure 80.A&B**]. Therefore, stable clones are most likely enriched for non-expanded naïve B cells.

The baseline isotype usage for stable clones was perturbed in sepsis, with significant increase in class-switched isotypes and depletion of IgM and IgD compared to health. This reflects the absolute percentage change in isotype usage for the total repertoire observed at each time-point. When comparing health and disease, expanding clones were significantly enriched for both IGHG1 usage and IGHG1 count in sepsis. IGHM usage (but not count)

was depleted consistent with repertoire-level isotype usage. However, there was no difference in usage for the remaining isotypes between health and sepsis. This is interesting given the significant perturbations seen across the majority of class-switched isotypes in stable clones. This suggests expanding clones in health are also driven by antigen encounter but without the need for biased IgG1 usage which as discussed previously in **4.3.3.1**, is linked to *S. pneumoniae* infection.

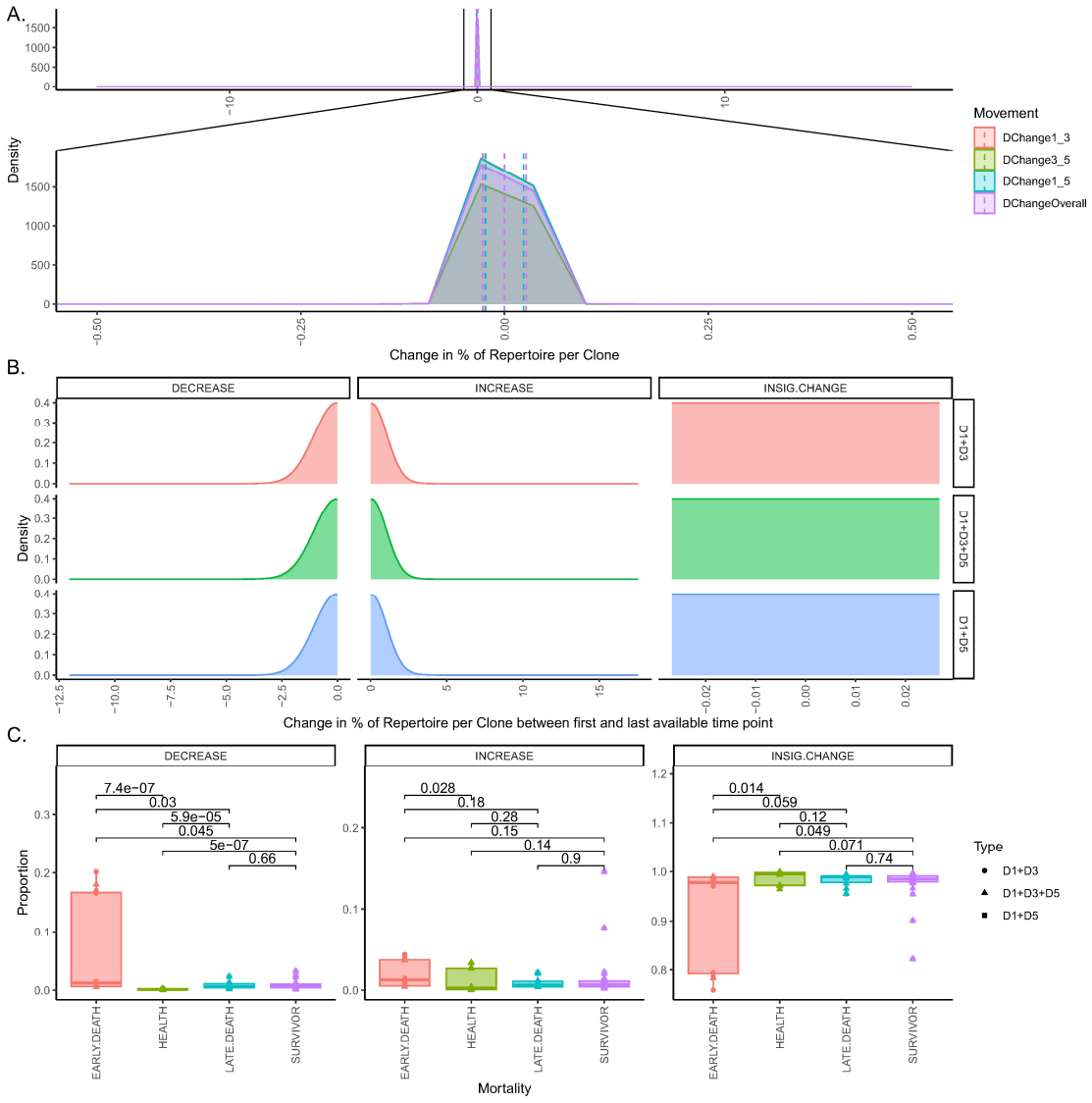


Figure 78. Investigating B cell clonal dynamics in sepsis.

A) Density plot showing the percentage change in repertoire for all B cell clones between the first and last time point. **B)** B cell clones were classified as increasing, decreasing or stable (insignificant change). **C)** The proportion of clones classified as each dynamic group, split by health and sepsis mortality group. P values for Wilcoxon tests are reported.

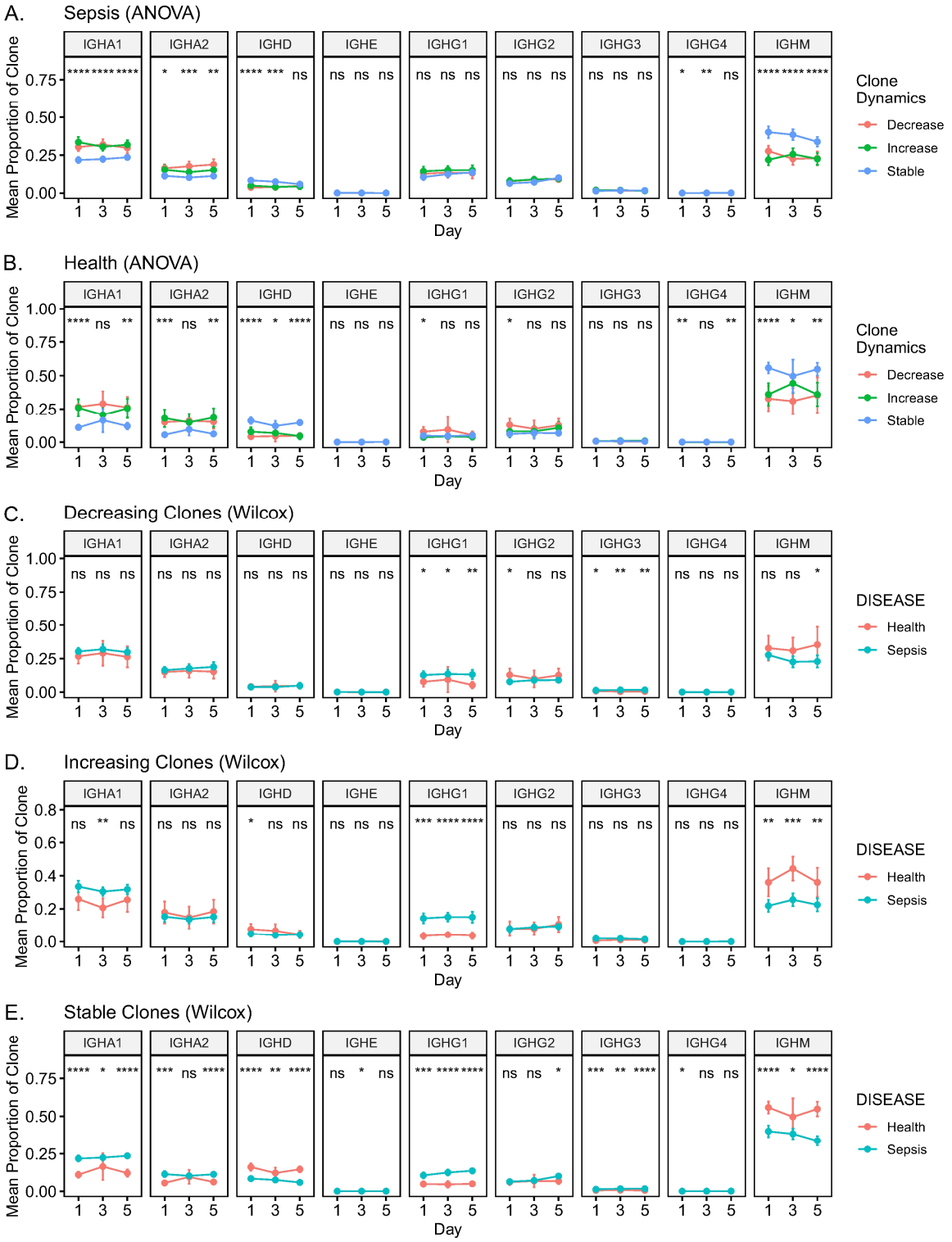


Figure 79. Longitudinal B cell clone isotype usage.

A, B) Isotype usage for **A)** sepsis and **B)** healthy clones grouped by clonal dynamics. Significance levels for ANOVA are shown. **C-E)** Comparison of clonal isotype usage between health and sepsis for **C)** decreasing clones, **D)** expanding clones and **E)** stable clones. Significance levels are shown for Wilcoxon tests.

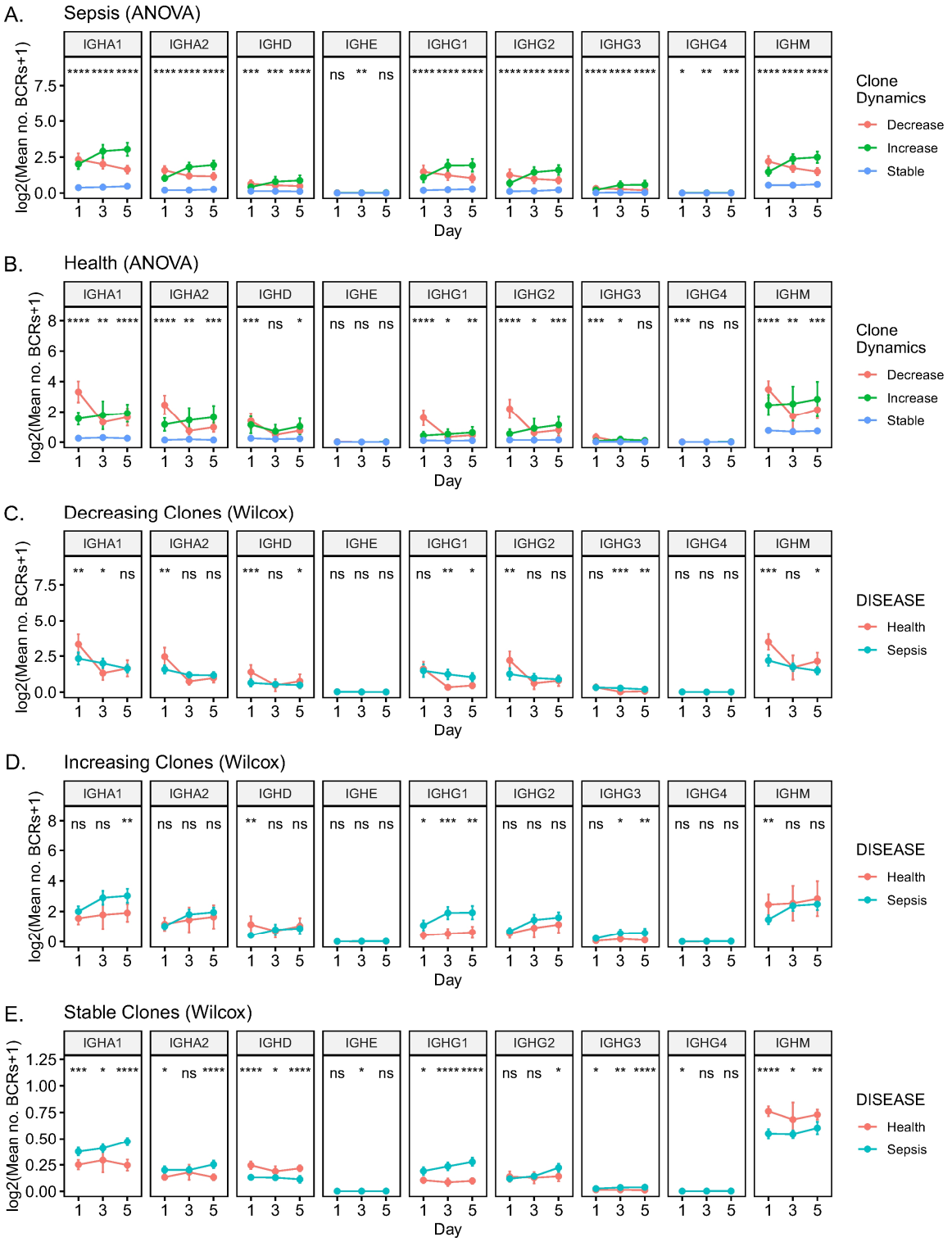


Figure 80. Longitudinal B cell clone isotype count.

A, B) Isotype counts for **A)** sepsis and **B)** healthy clones grouped by clonal dynamics. Significance levels for ANOVA are shown. **C-E)** Comparison of clonal isotype count between health and sepsis for **C)** decreasing clones, **D)** expanding clones and **E)** stable clones. Significance levels are shown for Wilcoxon tests.

4.4 Discussion

4.4.1 Unpaired IgH Chain Sequencing

To characterise the BCR repertoire in sepsis, I have performed NGS of the immunoglobulin heavy chain using bulk RNA, rather than the light chain. This is often the case in repertoire studies; the IgH chain is comprised of both V, D and J gene segments which enables more accurate detection of BCR clones. Additionally, as the IgH chain encodes the constant region it determines the downstream effector function (Yaari and Kleinstein, 2015). By comparison, the IgL chain repertoire is more limited, and usually dominated by a handful of IGLV and IGKV genes. Public clonotypes may account for as much as 60% of the human κ and λ repertoire (Collins and Watson, 2018). As the IgL and IgH pair to form the antibody binding site, IgL directly contributes to the specificity and binding properties of an antibody. The absence of paired-light chain sequencing data is one of the key limitations of this study, and would be required to determine the antibody specificity of interesting clones, for example through recombinant antibody expression.

Light chains κ and λ are also measurable as circulating free light chains (FLCs) separate from Ig molecules and are a by-product of excess light chain production during B cell antibody synthesis. They are secreted into the circulation where they are rapidly cleared by the kidneys (half-life of 2-6 hours) (Brebner and Stockley, 2013). In a large longitudinal study of the general population (n>15,000) non-clonal increase in FLC was a predictor of mortality, independent of age, sex and renal impairment (Dispenzieri et al., 2012). Elevated FLC levels are associated with low-grade inflammation, chronic immune stimulation, increasing age and renal dysfunction, which all predispose to sepsis. Interestingly, research has shown that sepsis patients have significantly more $CD19^{+}$ B cells expressing both κ and λ light chains, compared to healthy controls and this correlates with elevated free light chain

concentrations. Sepsis patients with high FLCλ also had a higher rate of mortality (Taccone et al., 2009, Shankar-Hari et al., 2017). Therefore, in conjunction with their short half-life FLC levels may provide a good marker of current disease activity.

The presence of dual CD19κλ is rare in health (~0.5%) and Shankar-Hari et al. (2017) suggest that the significant elevation seen in sepsis is due to allelic inclusion from impaired K deleting element activity (Xu, 2006a). They further hypothesise that elevated FLC may contribute to the prolonged neutrophil life span seen in sepsis via enhanced tyrosine phosphorylation. This is particularly interesting, because as discussed previously (**4.3.2.1**) neutrophil dysfunction is the likely cellular driver of the Sepsis Response Signature subgroups.

Emerging experimental technologies enable high-throughput paired IgH and IgL sequencing. Therefore, a more in-depth study could be performed in future to address some of the limitations of this analysis (DeKosky et al., 2013, DeKosky et al., 2015).

4.4.2 Sepsis and Accelerated Immunosenescence

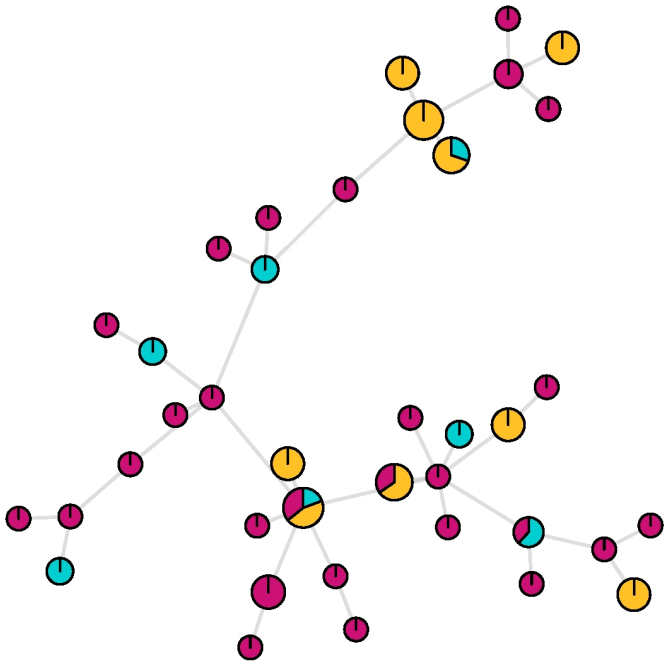
There were seven BCR modules (1, 5, 13, 16, 19, 21 and 31) that had a significant association with age, and these trends persisted when sepsis mortality was modelled independently of healthy controls. A common theme among age-associated modules was increased clonal expansion in response to increasing age. This is consistent with a publication by Gibson et al. (2009), which found that B cell diversity decreases in old age, is correlated with poor health status and was characterised by clonal expansion of B cells *in vivo* (de Mol et al., 2021, Weksler and Szabo, 2000). It has been hypothesised, that sepsis accelerates the development of immunosenescence (Lu et al., 2022). Interestingly, of the subset of age associated modules where sepsis was significant (13, 16, 19 and 31) both age and sepsis exert

their effect on BCR module score in the same direction. This suggests that sepsis may exaggerate the normal age-related changes to BCR repertoire, consistent with the idea of accelerated immunosenescence. During the five days sampled in this study, in many cases BCR module score did not return to baseline health, whether B cell clonal expansion returns to homeostasis during convalescence would be an interesting follow-up analysis to address this hypothesis.

4.5 Conclusion

In this chapter, I have used the rWGCNA pipeline to identify and understand perturbations to the BCR repertoire in a longitudinal sepsis cohort, in comparison to healthy controls. Notably, I provide evidence of a mortality-associated IgD^{+} plasma cell response in sepsis, likely driven by respiratory infection. In a novel analysis, I show how integration of BCR repertoire feature modules with additional experimental modalities, such as matched bulk RNA-seq data, can be used to understand regulation of the adaptive immune repertoire and have identified multiple associations with novel non-AIR genes which could function as potential biomarkers of the BCR repertoire.

Chapter 5. The TCR Repertoire in Sepsis



5 The TCR Repertoire in Sepsis

This chapter presents the results for the statistical analysis of the αβ T cell and γδ T cell receptor repertoire in sepsis, using the optimised repertoire weighted gene correlation network analysis (rWGCNA) pipeline. This chapter will also discuss high-level repertoire organisation, by integrating TCR and BCR repertoire modules.

5 The TCR Repertoire in Sepsis 274

5.1 Introduction 274

5.2 Chapter 5 Aims 280

5.3 Results 281

5.4 Discussion 356

5.5 Conclusion 360

***D* Appendices: The TCR Repertoire in Sepsis 421**

5.1 Introduction

This chapter focuses on the pathophysiology of the T cell receptor (TCR) repertoire in sepsis, beginning with a literature review. I employ the rWGCNA analysis pipeline developed in **Chapter 3** to analyse TCR repertoire data generated for the same cohort of sepsis patients and briefly discuss its performance.

In the second half of this chapter, I use the rWGCNA TCR modules to characterise and understand dynamic changes to the αβ and γδ TCR repertoire during acute sepsis, focusing on features associated with mortality, disease subtypes (SRS groups) and clinical variables. TCR modules are integrated with matched bulk RNA sequencing gene expression data and CIBERSORTx estimated cell fractions to understand regulation of the TCR repertoire.

Finally, I discuss the correlations between BCR and TCR modules, providing an understanding of the overall organisation of the adaptive immune repertoire within the context of sepsis

5.1.1 The TCR Repertoire in Sepsis

The study of the TCR repertoire in sepsis has faced similar limitations to those encountered for the BCR repertoire, including the use of out-dated molecular approaches and small, diverse patient cohorts. However, the role of T cells in sepsis has been studied more comprehensively than B cells (Ma et al., 2022). In the following sections, I review the scientific literature pertaining to the TCR repertoire in sepsis, focusing separately on αβ and γδ T cells.

5.1.1.1 Alpha-beta T Cell Receptor Repertoire

To understand the increased risk of nosocomial infections and death associated with sepsis, Venet et al. (2013) conducted a study to investigate combinatorial TCRβ repertoire diversity in 41 sepsis patients at days 1 and 7 post-onset of septic shock. The authors used a multiplex PCR approach to target the V-J region of the TCRβ chain, using genomic DNA (gDNA). Amplicons were size separated and their concentration was measured. Each individual V-J recombination was reported as a proportion of the total rearrangements. Results clearly show reduced TCRβ diversity at the onset of septic shock compared to healthy controls which increased (while remaining below healthy controls) over 7 days. Due to the rapid recovery of TCR diversity, the authors suggest that lymphocyte reconstitution after septic shock occurs through rapid mobilisation from lymphocyte peripheral stores rather than immune reaction-based oligoclonal expansion. No significant difference in TCR diversity was found between survivors and non-survivors, however reduced diversity was associated with an increased risk of developing nosocomial infections. No significant correlations were reported between diversity and either SOFA or SAPSII score, although a significant association was reported between TCR diversity and the number of HLA-DR expressing monocytes measured at day 1.

However, this study had several limitations. Firstly, the quantification of diversity was based on combinatorial diversity, within a single TCR chain. With the advent of NGS repertoire sequencing, more accurate quantification of TCR diversity and clonal expansion can be attained by including junctional diversity. Secondly, the study compared sepsis patients against normal values defined for healthy volunteers from a separate larger study, rather than using well-matched healthy volunteers. Covariates including age and sex were not incorporated in the statistical analysis. Finally, minimal information on infectious aetiology was provided, with the exception of the absence of infection caused by bacteria producing superantigens, limiting the reproducibility of these findings. However, a recent study by Tomino et al. (2017) which focused on patients with septic shock and latent cytomegalovirus infection reported similar findings, including decreased TCR diversity and monocyte HLA-DR expression in the early stages of septic shock. Measures normalised by day 7, except in patients who died, and was accompanied by persistent lymphopenia. Another study conducted in Taiwan by Cheng et al. (2021) used MiXCR to reconstruct TCR sequences from bulk RNA sequencing data for a small cohort of sepsis patients, and also reported a loss of TCR diversity.

In individuals receiving a liver transplant, sepsis is a leading cause of mortality in the post-transplant period. Interesting research conducted by Jones et al., 2021 showed that pre-transplant elevated TCRβ clonality was the only predictor of sepsis 3 months and 12 months after transplant. This suggests that repertoire perturbations could be a novel predictor of sepsis, occurring prior to the development of clinical symptoms. However, the increased risk of developing sepsis did not correspond to an increased incidence of mortality.

In contrast, Sun et al. (2017) sequenced the CDR3β repertoire of 20 neonates with sepsis at day 1 after diagnosis, and day 7 of treatment using gDNA, multiplex PCR and NGS. Results

were compared to the repertoire of 8 matched neonatal controls without infection. TCRβ diversity was generally lower in neonates with sepsis but the difference was not significant. Similarly, no significant differences were detected in V gene usage, however visualisation of principal components clearly separated sepsis, sepsis treatment and control groups; evidence of subtle changes to the repertoire which must be considered collectively. CDR3β sequences were classified by frequency, revealing an enrichment of rare clonotypes and absence of hyper-expanded clonotypes (more than 1%) in neonates with sepsis, suggesting limited repertoire focusing and expansion. The discrepancy between the results of this study and that of Venet et al. (2013) could be explained by the immature immune system in neonates characterised by reduced capacity for T cell expansion, the substantially smaller sample size and differences in causative pathogen.

To investigate long-term immunosuppression after sepsis, Ammer-Herrmenau et al. (2019) used a mouse model of peritoneal contamination and infection to study impairments in $CD4^+$ T cell response at 1 week, 1 month and 3.5 months after sepsis onset. Multiplex PCR and NGS was used to amplify the CDR3β repertoire, and results showed no global decrease in TCRβ diversity irrespective of treatment with IL-7 at any time point. However, mice that survived sepsis did show a reduced number of antigen specific T cells when challenged with fungal antigen attributed to the loss of particular T cell clones.

These studies provide a promising foundation for a more in-depth analysis of the αβ TCR repertoire in sepsis. Sequencing both the α and β chain will provide a more complete overview of disease specific perturbations, including quantification of semi-invariant T cells such as MAIT cells. Furthermore using a large, well characterised patient cohort with reduced heterogeneity may help to understand the discrepancies across results reported in the literature.

5.1.1.2 Gamma-delta T Cell Receptor Repertoire

Gamma-delta T cells are the most abundant epithelial-resident lymphocyte and are important in mucosal defence against pathogens. Yet the γδ TCR repertoire is the most understudied. Previous research has shown a pronounced decrease in the number of γδ T cells in sepsis, which decreases further when sepsis worsens and there is a corresponding association with mortality (Andreu-Ballester et al., 2013, Liao et al., 2017b). Furthermore, they exhibit reduced antigen presentation and an inability to promote efficient proliferation of $CD4^+$ αβ T cells (Yang et al., 2021). In a study of 30 sepsis patients conducted by Wang et al. (2020a), the number of Vδ1 T cells was significantly increased in sepsis, compared to healthy controls and corresponded to a decrease in Vδ2 T cells. Vδ1 T cells are immunosuppressive; interestingly, the proportion of Vδ1 T cells was positively correlated with SOFA score, and an elevated ratio of Vδ1 to Vδ2 T cells was associated with lower patient survival. These findings were replicated by Jin et al. (2017), who additionally found that the number of immunosuppressive Vδ1 T cells in peripheral blood was elevated in patients with septic shock compared to sepsis and positively correlated with lactic acid level and APACHE II score. Likewise, a switch in the proportions of Vδ1 and Vδ2 cells in sepsis compared to health was reported by Diosa-Toro et al. (2013) for a cohort of 26 patients with sepsis, of which ~25% had sepsis due to CAP and 25% due to soft tissue infection.

In a mouse model of CLP and acute lung injury, absolute numbers of γδ T cells were shown to progressively increase in the lungs during sepsis, reaching a 7-fold increase at day 7. This could be due to local proliferation, low cell death rate or recruitment from other sites. Although there was no histological evidence of lymphoid cell proliferation or apoptosis suggesting recruitment is more likely, which may partly explain the decrease in peripheral γδ T cell count frequently reported (Hirsh et al., 2004). In agreement with this hypothesis,

de Souza Costa et al. (2015), showed that γδ T lymphocytes (but not αβ T cells) migrated from the spleen to the lungs of CLP mice, and the lung resident population was enriched for Vγ4 T cells. Treatment with Anti-Vγ4 mAb resulted in more rapid death after disease onset, suggesting a protective role for this population. In a similar study, mice deficient in γδ T cells were shown to have reduced survival times and increased bacteraemia and tissue damage compared to wild-type controls, which was associated with a decrease in the number of peritoneal neutrophils and myeloid suppressor cells (Tschöp et al., 2008).

5.1.2 The Effect of Age and Sex on the TCR Repertoire

The natural process of biological ageing results in alterations to the TCR repertoire for both αβ and γδ T cells, which are important to consider during repertoire analysis. In αβ T cells immunosenescence is characterised by a linear decrease in lymphocyte count, TCR diversity and the percentage of naïve T cells. There is also an increase in clonal expansion of memory $CD8^+$ T cells and greater clonal overlap between longitudinal samples with advancing age. However, interestingly individuals at extreme old-age exhibit higher TCRβ diversity compared to the previous age group, suggesting an association between these parameters and longevity (Britanova et al., 2014, Yoshida et al., 2017, Egorov et al., 2018, Johnson et al., 2021).

Gamma-delta TCR repertoire diversity also decreases with age and the prevalence of γδ T cell subsets is skewed, as certain populations are more resilient to ageing than others (Xu et al., 2020a, Aliseychik et al., 2020). The prevalence of the Vδ2+ subset decreases after age 30, a process which is also influenced by sex and these cells show a reduced responsiveness to stimulation. By comparison the Vδ1+ population is maintained, or in some cases increased (Clark and Thomas, 2020). Males typically have more γδ T cells than females,

which is likely driven by an increase in the Vδ2+ subset. There is also an association between ethnicity and variations in Vδ1+ T cell frequency (Clark and Thomas, 2020).

Age-related alterations to the TCR repertoire are similar to sepsis disease perturbations. This may partly explain the discrepancies in the literature, particularly where cohorts have a large difference in mean age. Therefore, the choice of age and sex matched controls is of utmost importance (Xu et al., 2020b).

5.2 Chapter 5 Aims

This chapter aims to characterise the dynamic changes to the αβ and γδ TCR repertoire in sepsis using the cohort of patients introduced in **Chapter 3**. The main objectives are:

1. To provide a comprehensive overview of TCR repertoire quality control (QC) analyses.
2. To summarise the performance of the rWGCNA pipeline on TCR repertoire data.
3. To identify the most suitable statistical models for analysing the TCR repertoire modules using Akaike Information Criterion (AIC) model selection.
4. Using the statistical models from aim 3, to identify TCR modules significantly associated with:
    a. the presence of sepsis
    b. sepsis mortality
    c. Sepsis Response Signature (SRS and SRSq).
5. To identify TCR repertoire features that drive module score in disease-associated modules, using PC loading scores. Then to explore the biological relevance of these modules by investigating their correlations with clinical data.
6. To enhance understanding of the TCR modules from aim 3 by performing integrative analysis, examining the correlation between module score and:

i. Gene expression data obtained from matched bulk RNA sequencing, followed by gene pathway analysis using XGR.

ii. Estimated cell fractions derived from bulk RNA sequencing data using CIBERSORTx.

7. To explore high-level repertoire organisation, through correlation analyses between BCR and TCR modules.

5.3 Results

5.3.1 Summary of Study Cohort

TCR-α, β, γ and δ repertoire data was generated for the cohort of sepsis patients and healthy volunteers described in **Chapter 3**. RNA was reverse transcribed using either a combined αβ or γδ reaction. Then each chain was amplified separately using a multiplex PCR reaction. Amplicons were library-prepped, multiplexed and sequenced in batches throughout 2021-2022 using 300bp paired-end MiSeq sequencing [**Table 26**]. The experimental protocol underwent extensive optimisation specifically for this study, details can be found in **Chapter 2**. To detect batch effect, a technical replicate 'JR-1975-1003' was re-amplified and sequenced with each sequencing submission. Data pre-processing was performed in parallel using the same in-house pipeline as described for the BCR repertoire analysis. The subsequent section (**5.3.2**) provides a detailed account of the quality control analyses performed.

No. of Samples	TCR Chain	Project	Library	Sequencing Lane	Cohort	Date
59	α	P210054	1-8	4	Sepsis Cohort 1	06/04/2021
73	α	P210054	9-18	5	Sepsis Cohort 1	06/04/2021
45	α	P210425	19-24	10	Sepsis Cohort 2	18/11/2021
37	α	P220204	25-30	13	Healthy Volunteers	27/06/2022
30	β	P210054	1-4	2	Sepsis Cohort 1	29/03/2021
80	β	P210054	5-15	3	Sepsis Cohort 1	24/03/2021
22	β	P210054	16-18	4	Sepsis Cohort 1	06/04/2021
45	β	P210425	19-24	10	Sepsis Cohort 2	18/11/2021
37	β	P220204	25-30	13	Healthy Volunteers	27/06/2022
89	γ	P210353	1-12	6	Sepsis Cohort 1	07/09/2021
44	γ	P210353	12-18	7	Sepsis Cohort 1	07/09/2021
45	γ	P210425	19-24	9	Sepsis Cohort 2	10/11/2021
37	γ	P220161	25-30	12	Healthy Volunteers	13/05/2022
44	δ	P210353	1-6	7	Sepsis Cohort 1	07/09/2021
88	δ	P210353	7-18	8+8b*	Sepsis Cohort 1 (repeated)	06/09/2021+ 10/11/2021
45	δ	P210425	19-24	9	Sepsis Cohort 2	10/11/2021
37	δ	P220161	25-30	12	Healthy Volunteers	13/05/2022

Table 26. Sequencing timeline for the TCR repertoire libraries generated using multiplex PCR.

Due to recruitment timelines amplification and sequencing of healthy volunteer and sepsis samples was performed separately. Sequencing was performed by the Oxford Genomics Centre. TCRD lane 8 was re-sequenced (8b) due to low read depth, the corresponding library FASTQ files were merged.

5.3.2 Quality Control Analysis

5.3.2.1 Read Depth

The number of unique TCR sequences and the percentage of reads which passed filtering was compared across sequencing lanes and libraries to identify potential batch effect **[Figure 81, Table 27]**. Read depth QC measures were comparable within T cell type. However, on average 2/3 fewer unique TCR δ/γ VDJ sequences were detected compared to TCR α/β

VDJs. As γδ T cells constitute ~5% of circulating T cells this is a comparatively good capture rate (Allison and Garboczi, 2002, Santana and Esquivel-Guadarrama, 2006). The percentage of TCR α/β reads which passed filtering was double that of TCR δ/γ despite using many more V gene primers in the multiplex PCR reaction, increasing the opportunity for primer-dimer formation. This suggests that a limiting factor for TCR δ/γ repertoire amplification was input RNA quantity; 500ng of RNA was used for TCR δ/γ reactions compared to 150ng TCR α/β. This corresponds to a 3.3x increase in input RNA for a population that is ~19x less prevalent and may have been insufficient to achieve similar read depth. For TCR α/β the ~50% filtering pass-rate is similar to the BCR 60% pass-rate for PCR amplification and MiSeq, reported by Bashford-Rogers et al. (2014) and for the BCR data generated in **Chapter 3**. Like the BCR protocol, in future, double-sided size selection during library preparation may increase yield by reducing primer-dimers and off-target products.

The 6 samples from individuals UK17730209 and UK59580028 excluded from the BCR analysis due to significant comorbidities were also excluded from the TCR analysis. A small number of samples were excluded based on unique productive VDJ read depth less than 300 for TCR α and β and less than 50 for TCR γ and δ. Thresholds were chosen so that ideally less than 10% of samples were excluded. Of the 205 sepsis samples originally selected, after filtering there were 176 (85.9%), 190 (92.7%), 168 (82%), 185 (90.2%), TCR α, β, γ, δ samples remaining respectively.

Given the absence of SHM, there are typically fewer unique TCR sequences compared to BCR sequences, particularly for TCR γ/δ where the number of V genes is limited. During pre-processing, identical VDJ sequences with different UMIs (which may or may not come

from the same cell) are combined into a reduced number of unique TCR sequences. Hence a smaller subsample depth is required compared to the BCR repertoire.

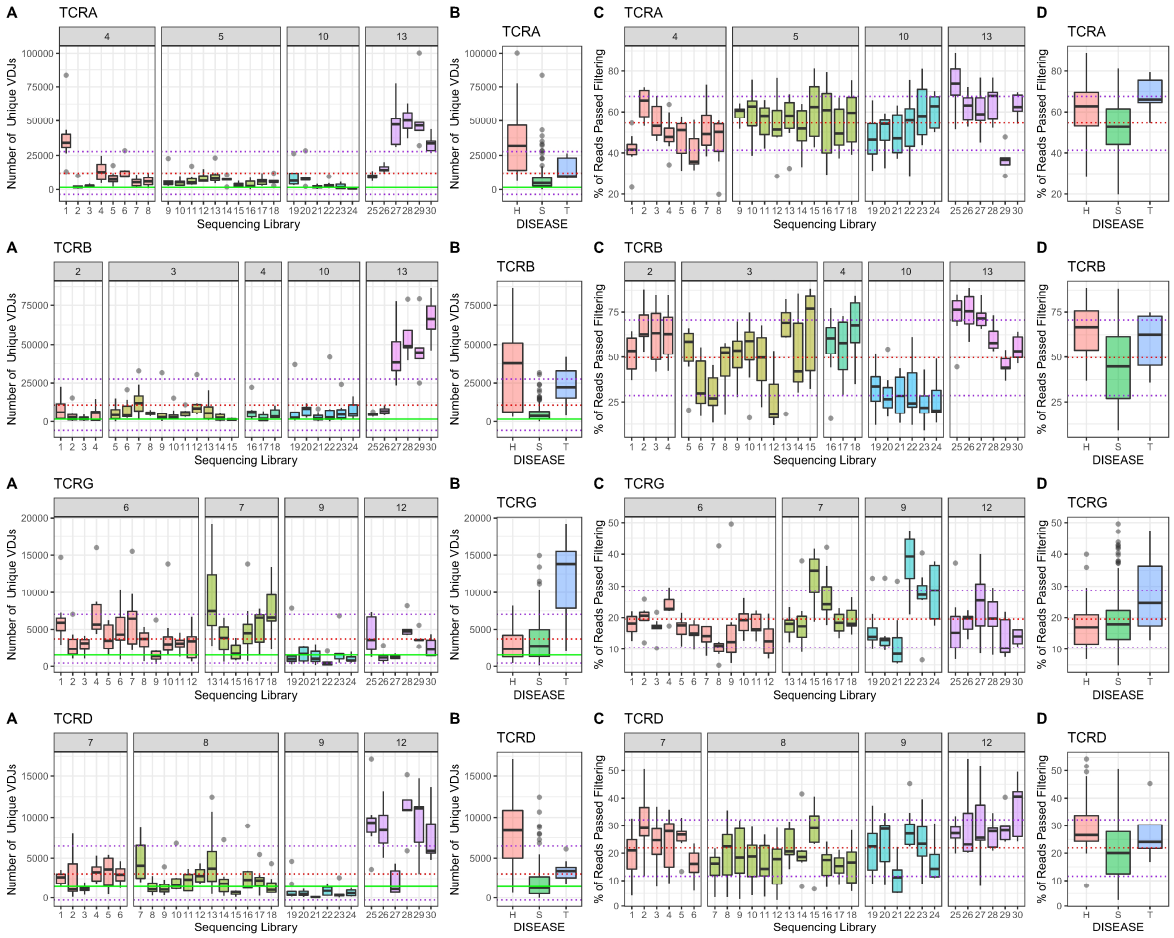


Figure 81. QC summary: TCRα, β, γ and δ repertoire sequencing depth.

A, B) The number of unique VDJs which passed the pre-processing pipeline grouped by library and sequencing lane **(A)** and disease status **(B)**. **C, D)** Percentage of VDJs which passed the pre-processing pipeline grouped as in A/B. Reads are excluded based on failure to map to known V or J gene segments and/or low quality (Phred score <30). The mean for the dataset is shown by the red line, purple lines are the mean ± 1 standard deviation. The horizontal green line is the 1500 read depth threshold used for BCR analysis as a comparison of read depth.

TCR	Lane	No. Unique VDJs Post Filtering			% Reads Passed Filtering		
		Mean	SD	NSD	Mean	SD	NSD
α	4	11,741.36	13,989.96	1.19	49.24	11.91	0.24
α	5	6,411.57	4,374.81	0.68	56.2	12.49	0.22
α	10	4,612.27	5,854.02	1.27	52.97	12.32	0.23
α	13	32,422.27	2,1430.13	0.66	61.28	13.78	0.22
β	2	5,184	5,800.39	1.12	60.74	13.98	0.23
β	3	6,603.69	7,013.03	1.06	47.54	20.4	0.43
β	4	4,995.5	4,987.45	1	58.37	18.29	0.31
β	10	6,647.96	8,621.69	1.3	28.98	12.83	0.44
β	13	33,521.81	27,202.18	0.81	65.5	13.94	0.21
γ	6	4,317.01	3,200.54	0.74	17.17	6.78	0.39
γ	7	5,599.55	4,164.26	0.74	22.36	8.48	0.38
γ	9	1,365	1,518.95	1.11	22.62	12.41	0.55
γ	12	2,964.68	2,061.86	0.7	18.17	8.22	0.45
δ	7	2,635.09	1,680.55	0.64	23.04	9.99	0.43
δ	8+8b	2,548.9	2,334.76	0.92	18.42	8.79	0.48
δ	9	665.93	842.39	1.26	21.1	10	0.47
δ	12	7,895.38	4,291.98	0.54	29.74	10.01	0.34
α	**All**	**12,106.26**	**15,580.95**	**1.29**	**54.45**	**13.12**	**0.24**
β	**All**	**10,958.44**	**16,600.19**	**1.51**	**49.62**	**21.02**	**0.42**
γ	**All**	**3,726.14**	**3,320.79**	**0.89**	**19.56**	**9.1**	**0.47**
δ	**All**	**3,095.06**	**3,387.18**	**1.09**	**21.89**	**10.27**	**0.47**

Table 27. QC summary statistics: TCRα, β, γ and δ repertoire sequencing depth.

Summary statistics include the mean, standard deviation (SD) and normalised standard deviation (NSD=SD/mean) for the number of unique VDJ sequences post filtering and the percentage of reads which passed filtering for each TCR (α, β, γ, δ) chain per sequencing lane and across all sequencing lanes (all).

5.3.2.2 V Gene Detection

In all TCR reactions, a low number of VDJ sequences from non-targeted TCR chains were detected. These reads predominantly mapped to the chain which pairs to form the TCR (*e.g.* α detection in the β reaction and γ detection in the δ reaction and vice versa). This can be attributed primarily to the use of combined TCRαβ and TCRγδ reverse transcription reactions prior to separate PCR amplification and degenerate PCR primers designed to target multiple V genes and alleles [**Figure 83**]. Downstream analysis was performed on productive reads from the targeted chain only.

Given the multiplex PCR reaction involved pooling up to 34 different forward V gene primers, there was an elevated risk of failing to combine all primers due to human error. This risk was greater than for the BCR repertoire, as only 6 forward primers were pooled.

To ensure all known V genes were successfully captured, total V gene count was calculated per TCR chain using sequences from all samples. The panel of detected V genes was compared to those included in reference file obtained from IMGT [**Figure 82**].

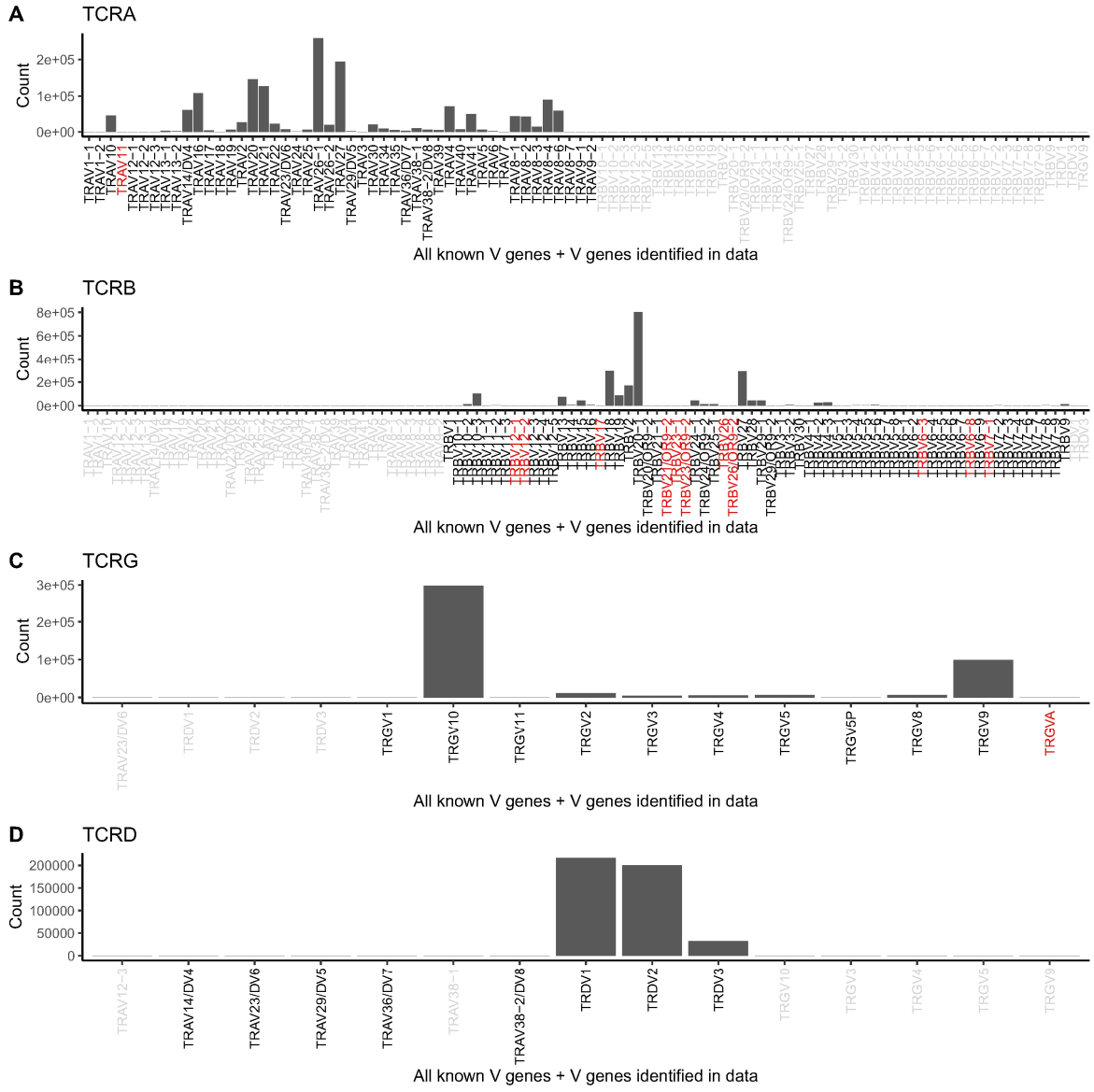


Figure 82. V gene detection in the TCR α (A), β (B), γ (C) and δ (D) repertoire datasets.

The total TCR V gene count across all samples is plotted. Genes in red are present in the reference file but were not detected in the repertoire data. V genes in light grey were detected at low levels in the dataset but do not belong to the targeted TCR chain.

5.3.2.2.1 TCRA and TCRB

All TRAV reference genes except for TRAV11 were successfully amplified. TRAV11 is a pseudogene and was not targeted with a PCR primer. The most frequently detected V gene was TRAV26-1.

In the TCRB dataset, several TRBV genes were absent. These were: TRBV12-1/2 (pseudogenes targeted by 'FT_TRBV12' degenerate primer), TRBV26, TRBV17, TRBV21, TRBV23-1, TRBV23, (pseudogenes/ORF for which no primer was designed) and TRBV7-1 (ORF targeted by 'FT_TRBV7' degenerate primer) [**Figure 82.B**]. Two functional genes, TRBV6-3 and TRBV6-8, were not detected. However, one degenerate primer ('FT_TRBV6') was designed to target all nine TRBV6 family genes, and of these 7/9 were detected including TRBV6 and TRBV9 which have identical primer binding sites to the two missing functional TRBV6 genes. Therefore, their absence is likely due to low prevalence in the UK population rather than problems with primer binding and amplification.

5.3.2.2.2 TCRG and the Implications of γ-αβ TCR Co-expression

TRGV, TRDV and γ/δ reverse transcription constant region primers were designed as part of this thesis and had not been previously validated. The only TRGV gene that was absent in our dataset was the pseudogene TRGVA, but no primer was included in the multiplex PCR reaction [**Figure 82.C**]. The most prevalent V genes in order of frequency were: TRGV10 (~70% of VDJs), TRGV9 (~30% of VDJs) and TRGV2. This presents an inconsistency with the literature. The most commonly used V gene associated with peripheral γδ T cells is TRGV9 (65-90% in adults), often in association with TRGVJP.

Exceptionally high capture of TRGV10 has been previously reported in RNA-based repertoire sequencing of patients with adult T cell leukaemia and lymphoma (Farmanbar et al., 2019). Surprisingly, the bias towards TRGV10 usage was restricted to VDJ sequences obtained from $CD4^+$ T cells and less than 1% of γδ T cells express CD4 (Garcillán et al., 2015). Similarly, a gDNA-based study found high TRGV10 usage in sequences from αβ T cells but not γδ T cells (Sherwood et al., 2011). A recent scRNAseq TCR repertoire study in rhesus macaques, reported co-expression of TRGV10 and TCRβ in around ~80% of γ-αβ T cells lineages, which comprised ~10% of T cells in PBMC samples (Walsh et al., 2021). Collectively these results are evidence of γ-αβ co-expression in αβ T cells after lineage commitment, and a bias towards TRGV10 usage. In humans, αβ T cells account for ~95% of all circulating T cells, compared to just 1-5% for γδ T cells (Cheng and Hu, 2017). Even if γ co-expression occurs in only a small proportion of αβ T cells, as reported in rhesus macaques (~10%), γ-αβ T cells would still outnumber true γδ T cells in humans by nearly double. This could explain why more than twice the number of TRGV10 VDJs compared to TRGV9 were detected in the repertoire data. Interestingly, TRGV10 has a non-consensus donor splice site in the first exon. This prevents splicing of the leader intron and causes termination of translation of the leader peptide (Sherwood et al., 2011, Antonacci et al., 2020, Aliseychik et al., 2020). Therefore, biased usage of TRGV10 could be one mechanism to prevent γ chain expression on αβ T cells (Sherwood et al., 2011).

Due to the uncertain origins of TRGV10 containing TCRs from unsorted populations, and their inability to form a properly folded TCR, all TRGV10 VDJs were excluded from downstream analysis, sacrificing considerable read depth. In future, the 'LEO_TRGV_3' primer which targets both TRGV10 alleles, may be excluded from the multiplex PCR reaction. Importantly, TRGV10-containing TCRs can be annotated as either productive or

non-productive by IMGT, as they are actively transcribed despite not generating a functioning protein. Our in-house computational pipeline groups reads based on functional annotation to ensure downstream analysis on productive reads only. Around 50% of TRGV10s were annotated as productive, therefore to exclude all TCRGV10 reads the IMGT output files were manually edited, setting all genes with an ORF annotation to 'non-productive'.

As TRGV9 was the second most prevalent TRGV gene detected (~30%), by excluding TRGV10 VDJs, we are confident that the remaining dataset captures sequences from true γδ T cells (Garcillán et al., 2015, Fichtner et al., 2020).

5.3.2.2.3 TCRD

All known TCRDV genes were detected, including the 5 TRAV/DV genes which can be used by both TCRα and TCRδ chains [**Figure 82.D**]. Peripheral γδ T cells are typically either TRDV1^{+} or TRDV2^{+} and these were found at the highest frequencies in our dataset (Cairo et al., 2010, Lee et al., 2019). Similar to the γ chain, δ chain V(D)J recombination occurs prior to T cell lineage commitment, however the arrangement of the delta locus within the alpha locus restricts cells to either δ or α expression; leading to silencing of the TCR-δ chain in αβ T cells (Hockett et al., 1988, Ziegler et al., 2014). Therefore, we can confidently attribute all TCRδ sequences to true γδ T cells.

Consistent with this conclusion, PCR amplification of the TCRδ chain was more challenging than TCRγ. The TCRδ PCR required more cycles and yielded less product with increased off-target amplicons, despite using the same volume of cDNA (20µl) from the combined reverse transcription reaction [**Figure 83**]. This could be explained by the increased RNA

available for the TCRγ reaction from γ-αβ co-expression, important in a reaction where RNA is a limiting factor.

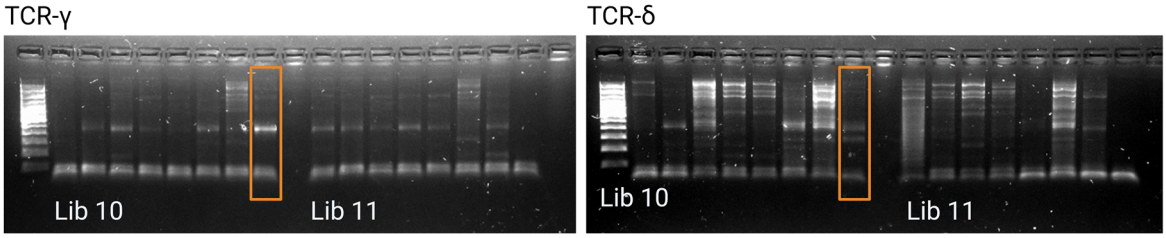


Figure 83. Agarose gel visualising the PCR product from TCRγ and TCRδ multiplex PCR amplification for the same 15 samples comprising libraries 10 and 11.

Input cDNA was generated from a combined γδ reverse transcription reaction. The TCRγ chain amplified more efficiently, requiring 32 PCR cycles instead of 33, despite using the same volume of input cDNA – particularly evident for library 11. Moreover, there was a reduced occurrence of off-target binding in the TCRγ reaction, despite both reactions utilising a similar number of primers. The technical replicate, serving as a positive control, is highlighted in orange. A RT negative control was included in library 11, well 8.

5.3.2.3 VDJ-C Gene Usage

5.3.2.3.1 Constant Region

To identify batch effect due primer amplification bias which can arise when using new primer mix aliquots, the percentage of the repertoire comprised of each constant gene across the technical replicates was assessed [**Figure 84**]. For both TCRB and TCRG there are two available constant genes, which were targeted by a single degenerate reverse transcription primer. TCRβ constant region usage was reproducible across technical replicates (~30% TRBC1, ~70% TRBC2). TCRγ constant usage was also reproducible (~80% TRGC1, ~20% TRGC2). However, a slight shift in proportions was observed for one technical replicate sequenced on lane 9. This replicate had a relatively low read depth, limiting its representativeness of the total repertoire.

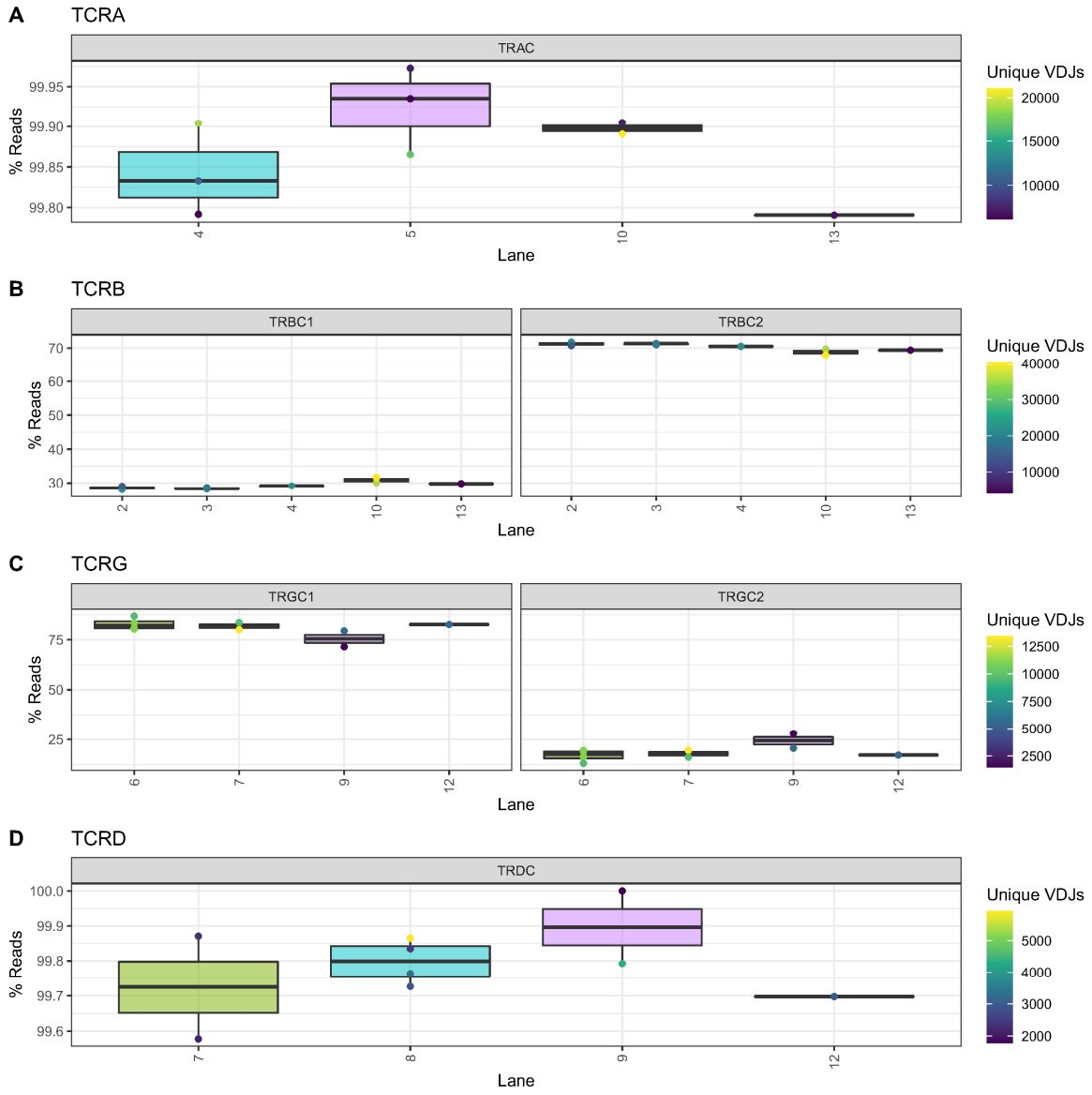


Figure 84. QC summary: TCR constant region usage for the technical replicates (healthy donor JR-1975-1003), grouped by sequencing lane.

Points are coloured by the total number of unique VDJ sequences for the replicate. Constant region usage may not total 100% due to low level detection of sequences from the chain that pairs to form the TCR receptor. The median, first and third quartile are shown and the whiskers represent 1.5x the interquartile range (IQR).

5.3.2.3.2 V and J Gene Usage

V and J gene usage was assessed across technical replicates, for each TCR chain [**Figure 85**]. The most commonly used V and J genes were consistent for all TCR chains. However, particularly evident in TCRA and TCRB is a systematic shift in V gene usage for technical

replicates included on the later sequencing lanes, which contained the healthy controls and sepsis cohort 2 (enriched for 7-day non-survivors). This is particularly evident for TRAV16, TRAV26-1, TRBV20-1 and TRBV28 [**Figure 85.A**]. To determine whether the systematic shift affected both technical and patient samples included on the same lane the Pearson Correlation Coefficient was calculated between the mean technical V gene usage per lane and mean sample V gene usage per lane. The strong positive correlation seen across all TCR chains indicates a systematic shift, affecting both the technical samples and patient samples. Examples are shown for TCRA and TCRD [**Figure 86.B** and **Figure 87.B**]. There was no overlap between healthy controls and sepsis patients on sequencing lanes, therefore this shift could result in the detection of a significant difference between health and disease which is driven by batch effect. Ideally, patient and control samples would have been interspersed on sequencing lanes, but as no healthy controls were recruited as part of GAinS and healthy volunteer recruitment was disrupted by the COVID-19 pandemic, this was not possible.

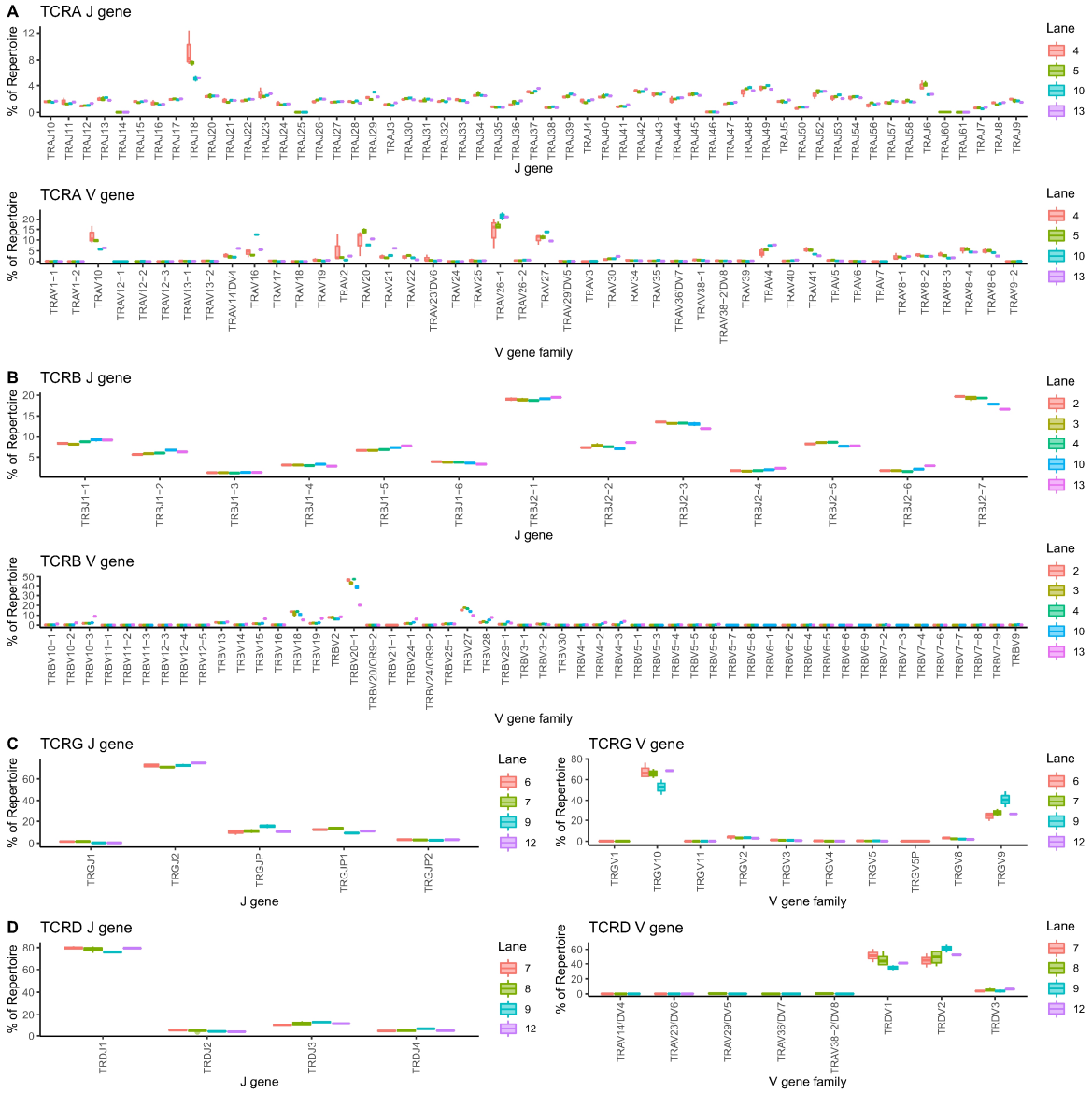


Figure 85. QC summary: the percentage of total unique VDJ TCR reads for each V and J gene for the technical replicates (healthy donor JR-1975-1003), grouped by sequencing lane.

A) TCRα **B)** TCRβ **C)** TCRγ and **D)** TCRδ technical replicates grouped by sequencing lane. Overall, technical replicate gene usage was similar across replicates for all chains. J gene usage was generally more reproducible than V gene usage. The median, first and third quartile are shown and the whiskers represent 1.5x the interquartile range (IQR).

5.3.2.3.3 V Gene Normalisation

In attempt to correct the V gene batch effect, the percentage of the repertoire explained by each V gene was normalised by subtracting the mean of the technical replicates included on the same sequencing lane. Where technical V gene usage was stable across lanes, all samples

are corrected similarly, and the distribution of the normalised data is much the same; as seen for TCRD **[Figure 87].** However, where technical V gene usage changes substantially across lanes, samples are corrected to different extents, altering their distribution. For example, for TRAV16 prior to normalisation, an elevation in V gene usage was observed for samples on lanes 10 (cohort 2, 7-day non-survivors) and 13 (healthy volunteers), coinciding with increased technical TRAV16 gene usage. It is unlikely to observe such markedly perturbed V gene usage between sepsis cohorts 1 and 2 due to biological factors, considering sepsis cohort 1 also includes non-survivors and was well-matched in terms of age, sex and ethnicity. However, a significant difference between health and disease is plausible. After normalisation, the mean TRAV16 usage for lane 10 (cohort 2) sepsis patients was similar to that of the sepsis samples sequenced on lanes 4 and 5. However, the elevated TRAV16 gene usage observed in health was preserved [**Figure 86.D**].

It should be noted that this normalisation approach is not without limitations. There will inevitably be variation in technical V gene usage due to re-amplification and sequencing. However, most sequencing lanes included multiple technical replicates, and normalising against their mean V gene usage helps account for stochastic variations in any individual replicate. In future studies, the inclusion of a positive control for each sequencing library could be beneficial. While this would slightly reduce sequencing depth for patient samples, it would enable library-specific or more accurate lane specific normalisation. Ideally, projects should aim to use a single primer aliquot for all samples, however given the potential for cross-contamination and in large studies this may not be feasible. J gene usage was not included as one of the summary metrics, but the same normalisation approach could be applied if necessary.

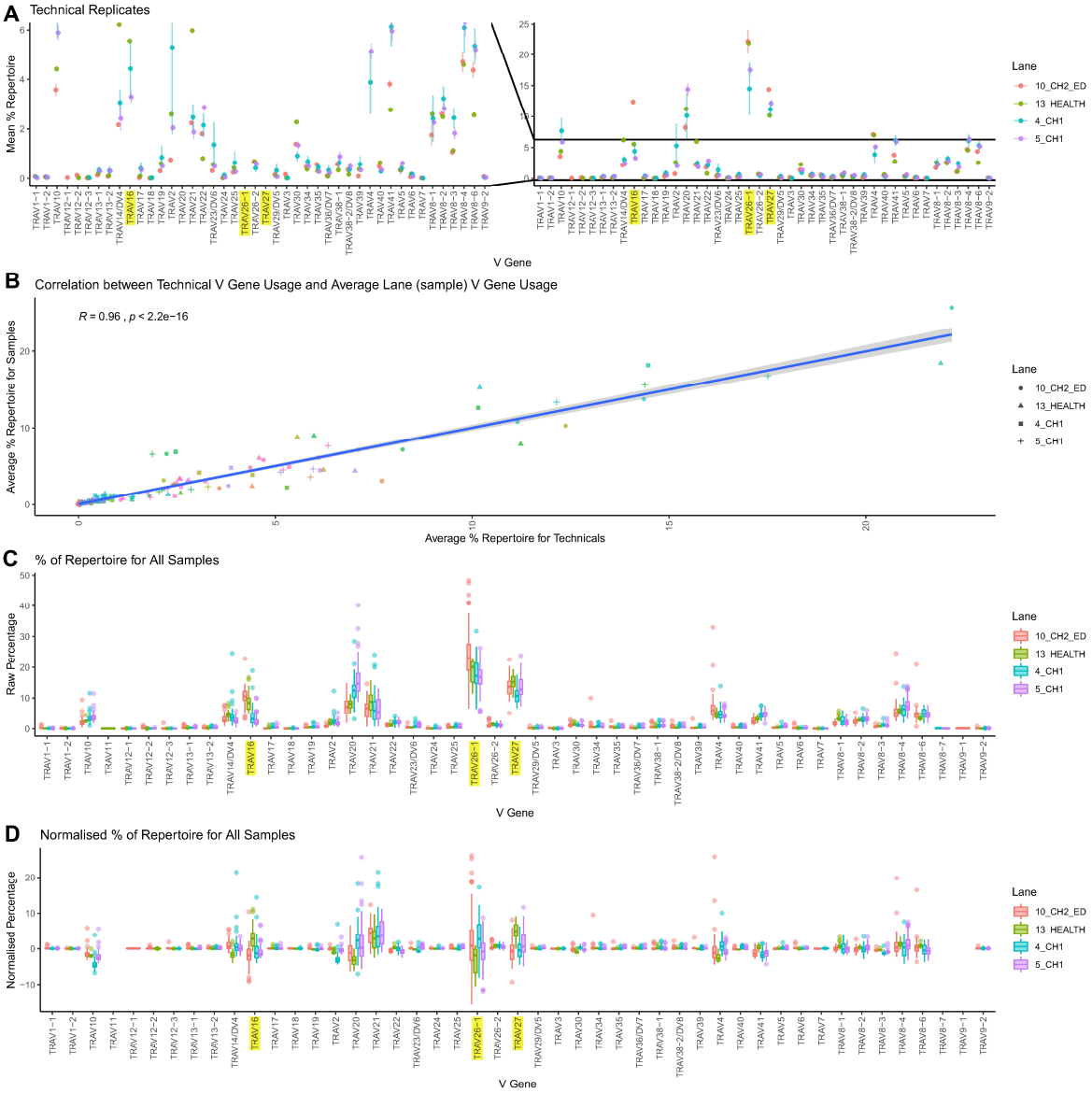


Figure 86. TRAV gene normalisation to account for batch effect detected in the technical replicates.

A) Mean technical TRAV gene usage across sequencing lanes, whiskers are the standard error of the mean. **B)** The correlation between mean technical TRAV gene usage and mean sample TRAV gene usage per lane, coloured by TRAV gene. A significant correlation suggests potential batch effect. **C)** Sample TRAV gene usage grouped by lane. **D)** Normalised sample TRAV gene usage per lane. Boxplots show the median, first and third quartiles and whiskers are 1.5x the interquartile range.

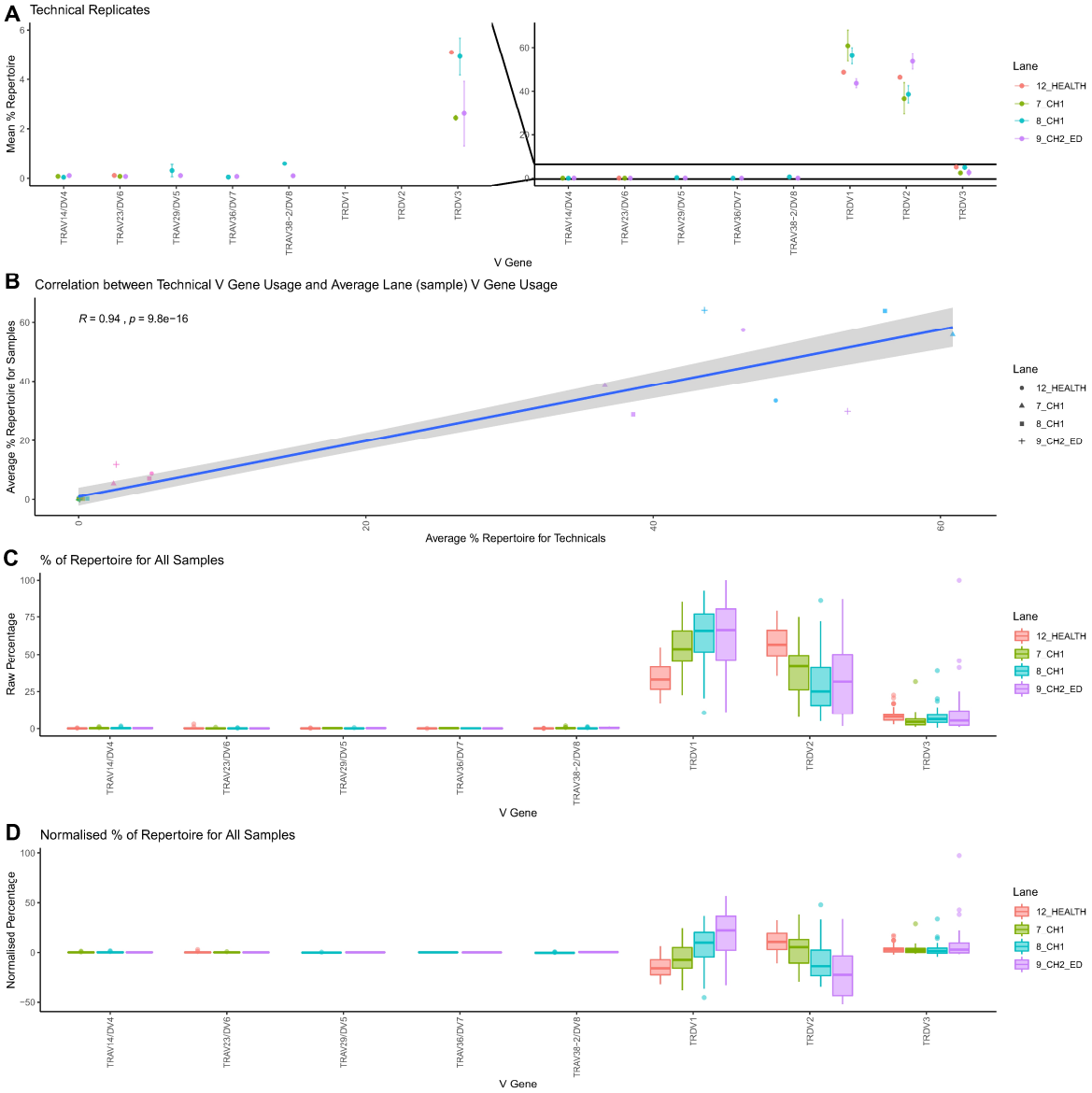


Figure 87. TRDV gene normalisation to account for batch effect seen in the technical replicates.

A) Mean technical TRDV gene usage across sequencing lanes, whiskers are the standard error of the mean. **B)** The correlation between mean technical TRDV gene usage and mean sample TRDV gene usage per lane, coloured by TRDV gene. A significant correlation suggests potential batch effect. **C)** Sample TRDV gene usage grouped by lane. **D)** Normalised sample TRDV gene usage per lane. Boxplots show the median, first and third quartiles and whiskers are 1.5x the interquartile range.

5.3.2.4 Index Hopping and Sample Contamination

To test for index hopping and sample contamination, the sequence overlap between all pairwise samples was calculated using the Jaccard index (JI). In contrast to the BCR, TCRs do not undergo SHM, so it was not surprising to observe that the JI between non-related samples was generally higher for TCRs [**Figure 88**]. Both TCRB and TCRD chains consist of V, D, and J gene segments, and for these chains, longitudinally related samples showed a slight increase in the JI [**Figure 88.B&D**]. This was not observed for TCR α and γ, which are comprised of V and J gene segments only [**Figure 88.A&C**]. There was some evidence of either low-level index hopping or contamination, but the reduced diversity of the TCR repertoire may not provide sufficient resolution to accurately detect this signal, compared to the BCR repertoire.

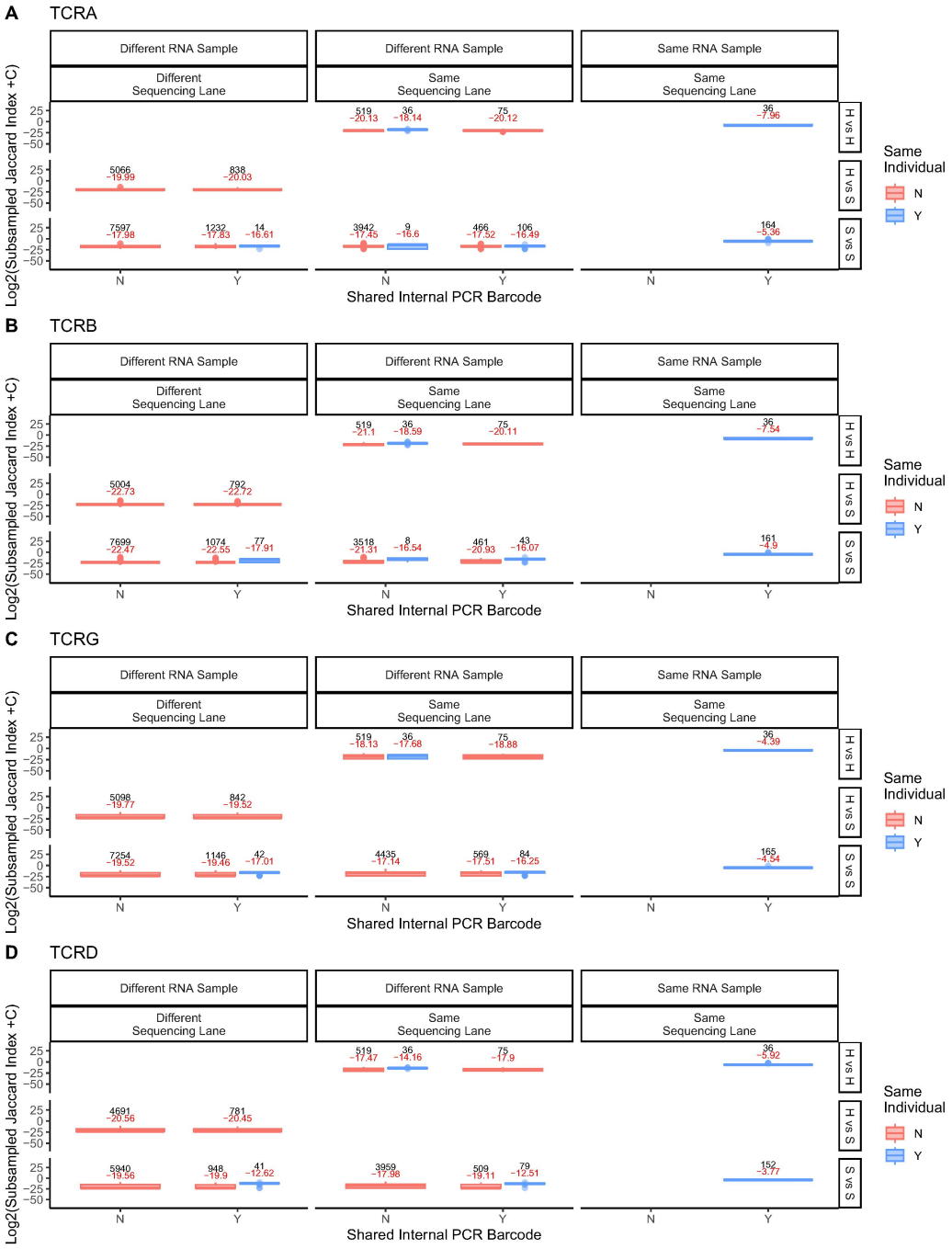


Figure 88. QC summary: TCR sequence overlap.

The Jaccard index was calculated to investigate TCR repertoire sharing between samples for each TCR chain. **A)** TCRα **B)** TCRβ **C)** TCRγ and **D)** TCRδ. The Jaccard index was calculated on a subsample of the repertoire to account for differences in read depth. The mean of 10,000 repeats was calculated. A small constant 'c' (c $=1/2\times$ minimum of the Jaccard index for which the index $\neq 0$) was added to avoid $\log_2(0)$ error. Sample overlaps are grouped according to whether samples were run on the same sequencing lane, had the same internal PCR barcode and whether they came from the same individual. As a reference scale for TCR overlap between samples, the Jaccard index was calculated on 2 independent down-samples samples of the same repertoire. The number of pairwise comparisons are shown in black; large differences in group size prevented statistical analysis. The mean for each group is reported to 2 decimal places in red. Technical replicates were excluded from this analysis. Boxplots show the median, first and third quartiles and whiskers are 1.5x the interquartile range.

5.3.2.5 Rarefaction Analysis

Rarefaction analysis was conducted to ascertain whether TCR read depth was sufficient to estimate VDJ diversity without normalisation [**Figure 89.A**]. Similar to the BCR repertoire, sequencing saturation was not reached for any TCR chain. Therefore, TCR measures were calculated as the mean of repeated subsamples, following best practices (Shugay et al., 2015, Chakraborty et al., 2022, Nakanishi et al., 2016). Although no significant T cell clonal expansion was detected in individual UK59580028 (leukaemia) and UK17730209 (metastatic tumour), these samples were still excluded from further analysis.

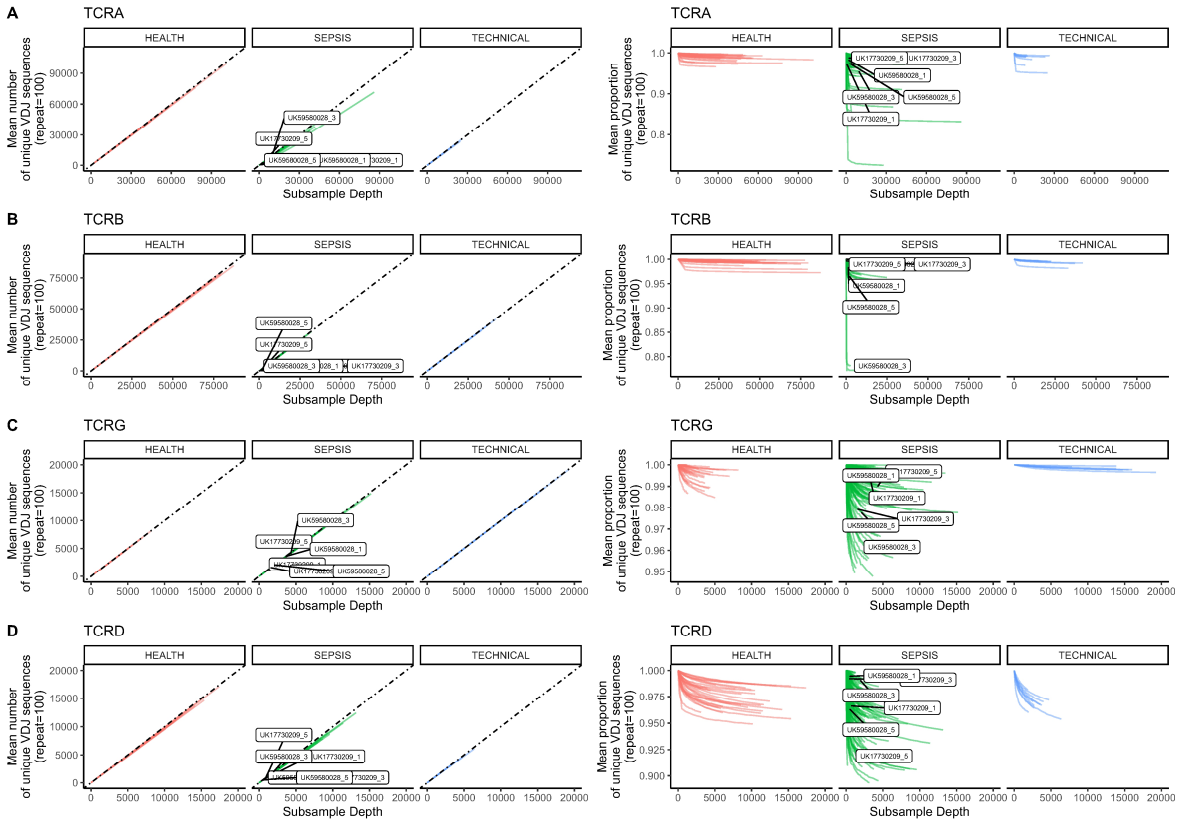


Figure 89. TCR rarefaction analysis.

A) TCRα **B)** TCRβ **C)** TCRγ and **D)** TCRδ, grouped by healthy donors, sepsis patients and the technical replicate. **Left)** The mean number of unique TCRs in a subsample of the repertoire, plotted against the size of the subsample for 100 repeats per sample and 20 subsamples of increasing size. The black dotted line represents a one-to-one relationship. **Right)** The mean proportion of unique TCRs in a subsample of the repertoire for 100 repeats, is plotted against subsample size per sample. Individuals UK17730209 and UK59580028 are marked as these samples were excluded due to significant comorbidities, although significant T cell clonal expansion was not detected.

5.3.2.6 Quality Control Analysis Summary

In summary, QC metrics were compared across technical replicates, and results suggest minimal batch effect between sequencing lanes and libraries in the TCR repertoire datasets. However, a shift in V gene usage was observed across sequencing lanes, which prompted the inclusion of a normalisation stage. Downstream analysis was conducted on each dataset without further batch correction.

Validation of primer sets revealed the unique challenges of TCR γ repertoire analysis using non-sorted populations due to γ-αβ co-expression. This analysis provides valuable guidelines to the wider scientific community as this issue has not been addressed in the existing literature.

TCR sequences were subsequently annotated using IMGT HighV-QUEST. Output files from IMGT were downloaded and parsed in the later stages of the pre-processing pipeline to calculate repertoire summary measures. However, fewer summary measures were computed for the TCR repertoires compared to the BCR repertoire as calculating measures per constant region is not necessary.

5.3.3 Evaluating the Performance of rWGCNA Analysis

Feature matrices were merged according to T cell type, only samples with data for both chains were retained. There were 158, and 172 samples for γδ and αβ respectively. The merged TCRAB and TCRGD feature matrices were separately input into the custom rWGCNA pipeline. Each stage of the methodology was optimised for each T cell type and the results are summarised in **Table 26** and sections **5.3.3.1-5.3.3.4**.

Stage	TCR αβ	TCR γδ
Missing Data Imputation	• **MissForest**. • There was no clear optimal method, however additive regression (Hmisc) performed particularly poorly. • No TCR features were excluded due to poor imputation performance.	• **MissForest.** • There was no clear optimal method • No TCR features were excluded due to poor imputation performance.
Longitudinal Correlation Network Generation	**All samples used.** Correlation between all days and: ■ **day 1:** R=0.92 p<2.2e−16 ■ **day 3:** R=0.82 p<2.2e−16 ■ **day 5:** R=0.82 p<2.2e−16	**All samples used.** Correlation between all days and: ■ **day 1:** R=0.95 p<2.2e−16 ■ **day 3:** R=0.94 p<2.2e−16 ■ **day 5:** R=0.92 p<2.2e−16
Linkage Method	**Ward's D2**	**Ward's D2**
Cluster Detection Method	Non-stable estimates using fixed-height cutting. Sensitive to maximum cluster parameter. Instead used **Dynamic Tree Cut**.	Non-stable estimates using fixed-height cutting. Sensitive to maximum cluster parameter. Instead used **Dynamic Tree Cut.**
Optimal Cluster Number	**17**	**12**

Table 28. A summary of rWGCNA TCR specific optimisations.

The rWGCNA pipeline was run on the combined TCRAB and TCRGD matrices and at each stage the optimal method was identified. In the majority of cases the optimal methods for the TCR repertoire were the same as for the BCR repertoire.

5.3.3.1 Missing Data Imputation

The optimal missing data imputation method was determined for each TCR repertoire separately, but no clear best method was identified for either [**Figure D. 1**, **Table 28**]. Reduced data complexity may explain why imputation methods performed similarly.

The final TCRAB feature matrix had no missing data, but there were two features with missing data in the TCRGD matrix: the percentage of the repertoire explained by the top 50 clones for both TCRG and TCRD. Individuals with less than 50 total clones will have an 'NA' which could result from low read depth or loss of diversity and increased clonal expansion. TCRGD missing data was imputed using missForest. There were no significant change in feature mean and the absolute percentage change in means was below 5%. This

work highlights the successful use of missForest for imputing both missing BCR and TCR repertoire data [**Figure D. 2**].

5.3.3.2 Distance Matrix

TCR repertoire feature correlation networks were calculated using all time points and separately for time point 1, 3 and 5 [**Figure D. 3**]. The TCRAB network derived using all time points was most similar to the network derived using day 1 samples [**Table 28**]. By comparison, the TCRGD network derived using all time points was equally well correlated to the networks derived from each independent time point [**Table 28**]. After filtering, the number of TCRGD samples was roughly equal for day 1, 3, and 5 (n=59, 52, and 47 out of 158 samples, respectively) while for TCRAB, the numbers were 73, 52, and 47 out of 172 samples, demonstrating the impact of an unbalanced paired design on longitudinal network generation. This is challenging to avoid in longitudinal studies truncated by death, and the equal numbers of TCRGD samples at each time point after filtering arose purely by chance.

5.3.3.3 Cluster Detection

Features were clustered using agglomerative hierarchical clustering with Ward's linkage method [**Table 28**, **Figure D. 4**]. As observed for the BCR repertoire, fixed-height cutting was heavily influenced by user parameters and the optimal number of clusters was unstable [**Figure D. 5**, **Figure D. 6**]. In contrast, Dynamic Hybrid Tree Cut produced a stable optimal number of clusters using a minimum cluster size of 3 and 5 for the TCRGD and TCRAB feature matrix respectively. Given the performance of DHTC across both BCR and TCR repertoire feature matrices, I recommend its use in future analysis. The optimal number of clusters for the TCRGD feature matrix was 12 and for TCRAB it was 17 [**Figure D. 7**, **Figure D. 8**].

5.3.3.4 Principal Component Analysis

PCA was performed on each cluster, using the imputed, scaled TCR repertoire feature matrix [**Figure D. 9**]. For TCRAB, more than 50% of the total cluster variance was captured by PC1 in one third of modules. For TCRGD ~60% of modules captured ≥50% of the cluster variance. However, similarly to the BCR repertoire, PCA performed poorly on modules containing V genes, which comprised the majority of TCRAB clusters. The within-cluster PC1 loading scores were less variable compared to the BCR repertoire; likely due to the small number of features in TCR clusters [**Figure D. 10**]. Downstream biological interpretation is therefore less challenging.

5.3.4 Statistical Analysis of TCR Repertoire Modules in Sepsis

In total, 17 TCRAB and 12 TCRGD repertoire feature modules were identified using the rWGCNA pipeline [**Table 29**, **Table 30**]. The following sections of this chapter will focus on the exploration of module score in relation to disease status, sepsis subtypes and clinical variables. This analysis aims to provide insights into the pathophysiology of the TCR repertoire in acute sepsis.

Module	Number of Features	Summary	TCR Alpha	TCR Beta
1	14	Clonality/ Clonal Diversification CDR3 Charge V Gene Usage		
2	13	Normalised V Gene Usage		
3	12	Normalised V Gene Usage		
4	11	Clonality/Clonal Diversification		
5	11	Cluster/Vertex Rényi		
		V Gene Usage		
6	10	V Gene Usage		
7	10	Proportion Unproductive		
		Proportion NKT/MAIT		
		V Gene Usage		
8	9	V Gene Usage		
9	8	CDR3 Length Bimodality/Kurtosis/Skewness		
		CDR3 Charge		
10	7	CDR3 Length Bimodality/Kurtosis/Skewness		
11	6	Hydrophobicity kurtosis/bimodality and proportion productive		
12	6	V Gene Usage		
13	5	V Gene Usage		
14	5	V Gene Usage		
15	5	V Gene Usage		
16	5	V Gene Usage		
17	5	CDR3 Length		
		CDR3 Hydrophobicity Distribution		
		V Gene Usage		

Table 29. TCRAB module summaries annotated according to TCR α or β chain.

For each module the constituent repertoire features are recorded and annotated by TCR chain (alpha or beta) where relevant.

Module	Number of Features	Summary	TCR Gamma	TCR Delta
1	7	Clonality		
2	7	Clonality		
3	7	Clonality		
		Proportion Productive		
4	6	Proportion Productive		
		CDR3 Length and Distribution		
		V Gene Usage **DV2**		
5	5	V Gene Usage: **GV3/4/5, DV4, DV7**		
6	5	V Gene Usage: **DV1**		
		CDR3 Charge		
		CDR3 Length		
7	5	CDR3 Hydrophobicity Distribution		
		V Gene Usage: **DV3**		
8	4	V Gene Usage and proportion unproductive		
9	4	Charge		
		Hydrophobicity Distribution		
		V gene Usage: **GV9**		
10	4	Clonal Expansion		
11	4	Clonality (Rényi)		
12	4	CDR3 Hydrophobicity		
		V Gene Usage: **GV8, GV2**		

Table 30. TCRGD module summaries annotated according to TCR γ or δ chain.

For each module the constituent repertoire features are recorded and annotated by TCR chain (gamma or delta) where relevant.

5.3.4.1 Model Selection

AIC-based model selection was conducted for each TCR module, using the same candidate models as described in **Chapter 4. Table 23.** In the majority of TCRAB and TCRGD modules, the inclusion of disease status and time point main effect terms improved model fit [**Figure 90**, **Table D. 1**]. However, the distribution of AICc scores revealed that the optimal models did not consistently perform significantly better than other models (ΔAICc <2) [**Figure D. 11**].

Our initial hypotheses were the same as the for the BCR repertoire: αβ and γδ TCR repertoires would undergo significant alterations in sepsis, with potential time-dependent changes in sepsis but not in health. Optimal models for TCRAB modules 2 and 8, as well as TCRGD modules 2 and 7, included the disease status by time point interaction term (D*T) which supports our hypotheses and justifies the inclusion of the interaction effect in the models. Therefore, analysis proceeded using the most complex interaction model which included all covariates and a sample ID random effect term (D*TASC).

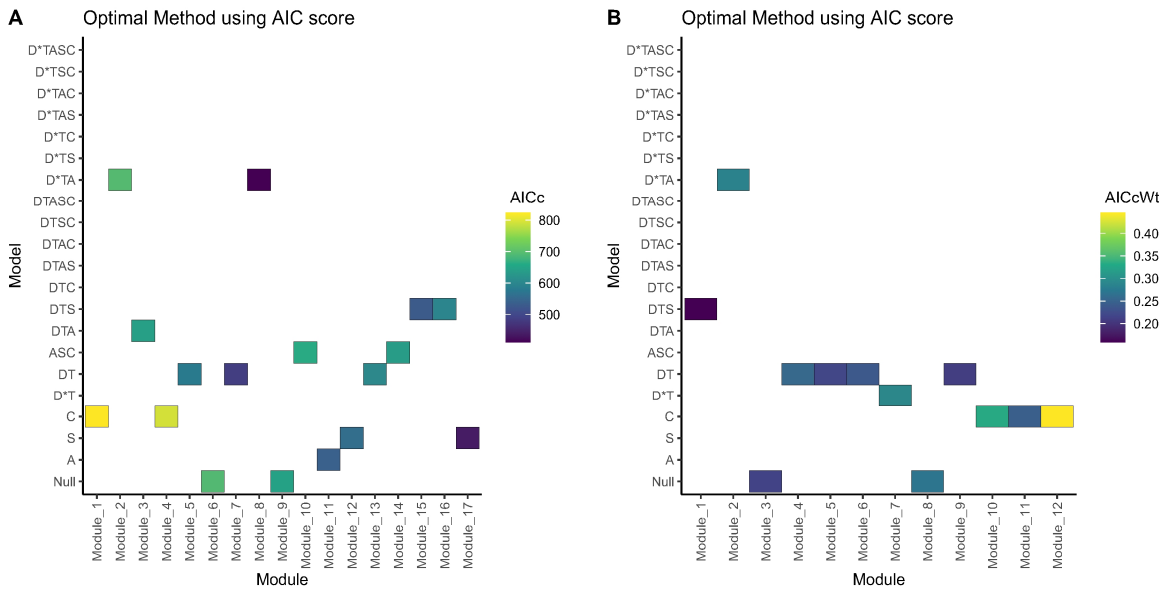


Figure 90. Optimal statistical model determined by AIC model selection.

Optimal linear mixed effect model from **Table 23** for **A)** TCRAB modules and **B)** TCRGD modules. Where models had the same AICc score the best was selected based on I) inclusion of the interaction term or II) the fewest predictors.

5.3.4.2 Health versus Sepsis

I applied the D*TASC linear mixed-effects (LME) model to each of the TCR modules separately to characterise sepsis-specific perturbations in the TCR repertoire. ANOVA was used to test for a significant improvement in model performance compared to the null (model: N) or covariate only model (model: ASC). In cases where a significant interaction between time point and disease status was observed, simple effects tests were performed to investigate the impact of time in sepsis or health (model: TASC) and the effect of disease status at day 1, 3 and 5 on module scores (model: DASC). Partial residual plots were used to visualise the interaction between health and disease for all modules controlling for covariates.

5.3.4.2.1 TCRAB

Out of the 17 TCRAB repertoire modules, six (2, 3, 5, 8, 13, 15) exhibited a significant difference between health and sepsis ($p<0.05$), with two modules (2 and 8) showing a significant disease by time interaction [**Figure 91**, **Table D. 2**]. All models explained significantly more variance than both the null and a covariate only model ($p<0.05$). Modules 2 and 8, showed a significant change in module score over time in health, as seen in the partial residual plots [**Figure 91.B&C**, **Figure 92**, **Table D. 3, Table D. 4**].

Instability in module score in health is likely due to differences in sample read depth, as of the 6 libraries containing healthy volunteer samples, libraries 27-30 had a substantially higher read depth (for both TCRB and TCRA) compared to libraries 25 and 26 [**Figure 20**]**.** This will be discussed further in **5.4.1**. For the two modules which were unstable in health, I will refrain from discussing temporal dynamics in sepsis in section **5.3.5** as further validation is needed. However, in both these modules, the substantial difference in module score between health and sepsis surpasses the temporal variability within each group, indicating genuine disease-specific perturbations in the repertoire. A detailed discussion of module biology will be presented in section **5.3.5.**

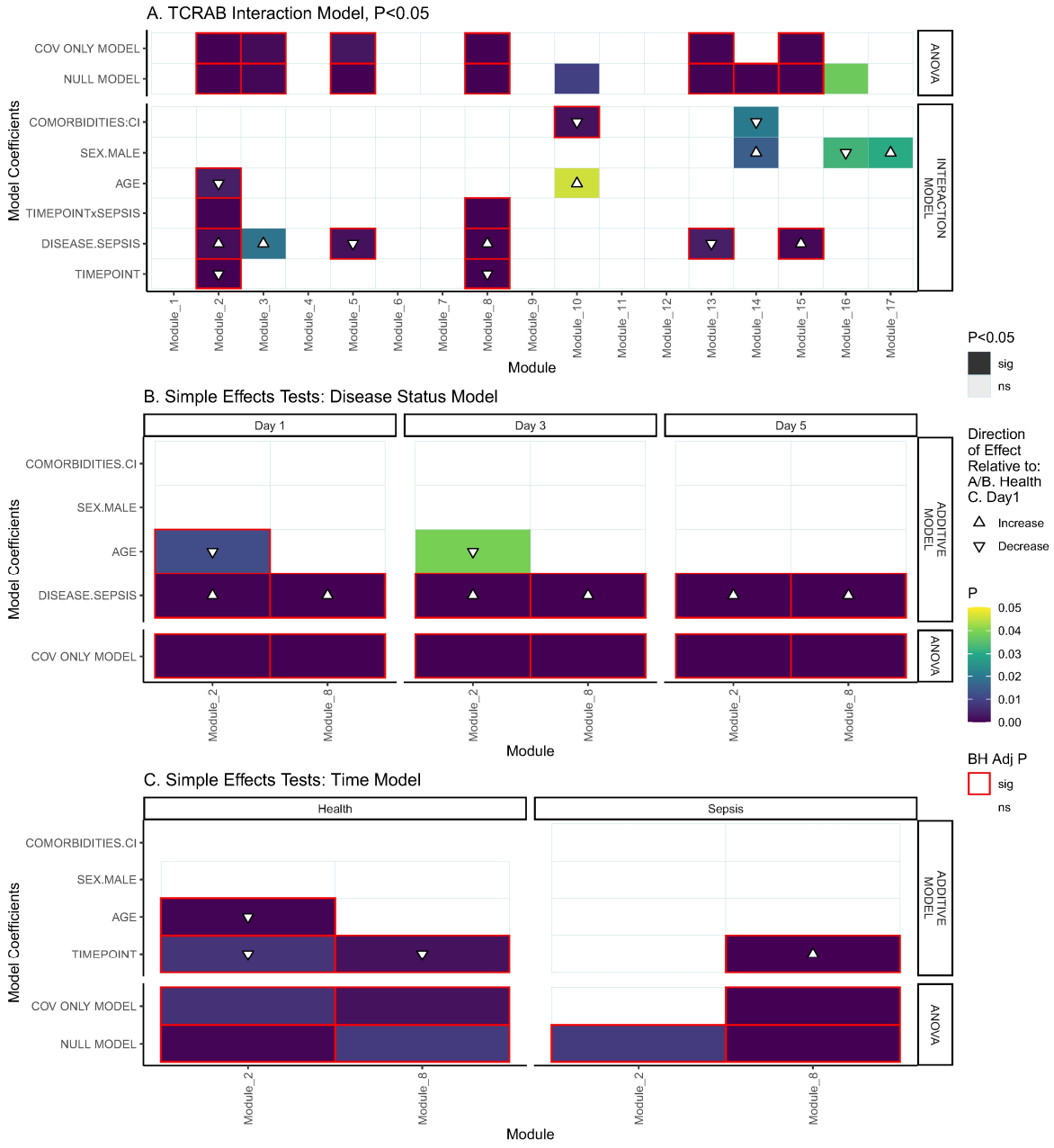


Figure 91. Linear mixed effect models investigating the effect of time and disease status on TCRAB module score.

A) The D*TASC model was applied to each module separately and significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the interaction model explained significantly more variance than either a null model or covariate only model. **B-C)**. Simple effects tests were conducted on modules with a significant interaction between time point and disease status and results are shown for: **B)** the effect of disease status (health or sepsis) at days 1, 3 and 5 and **C)** the effect of time on module score for either health or disease.

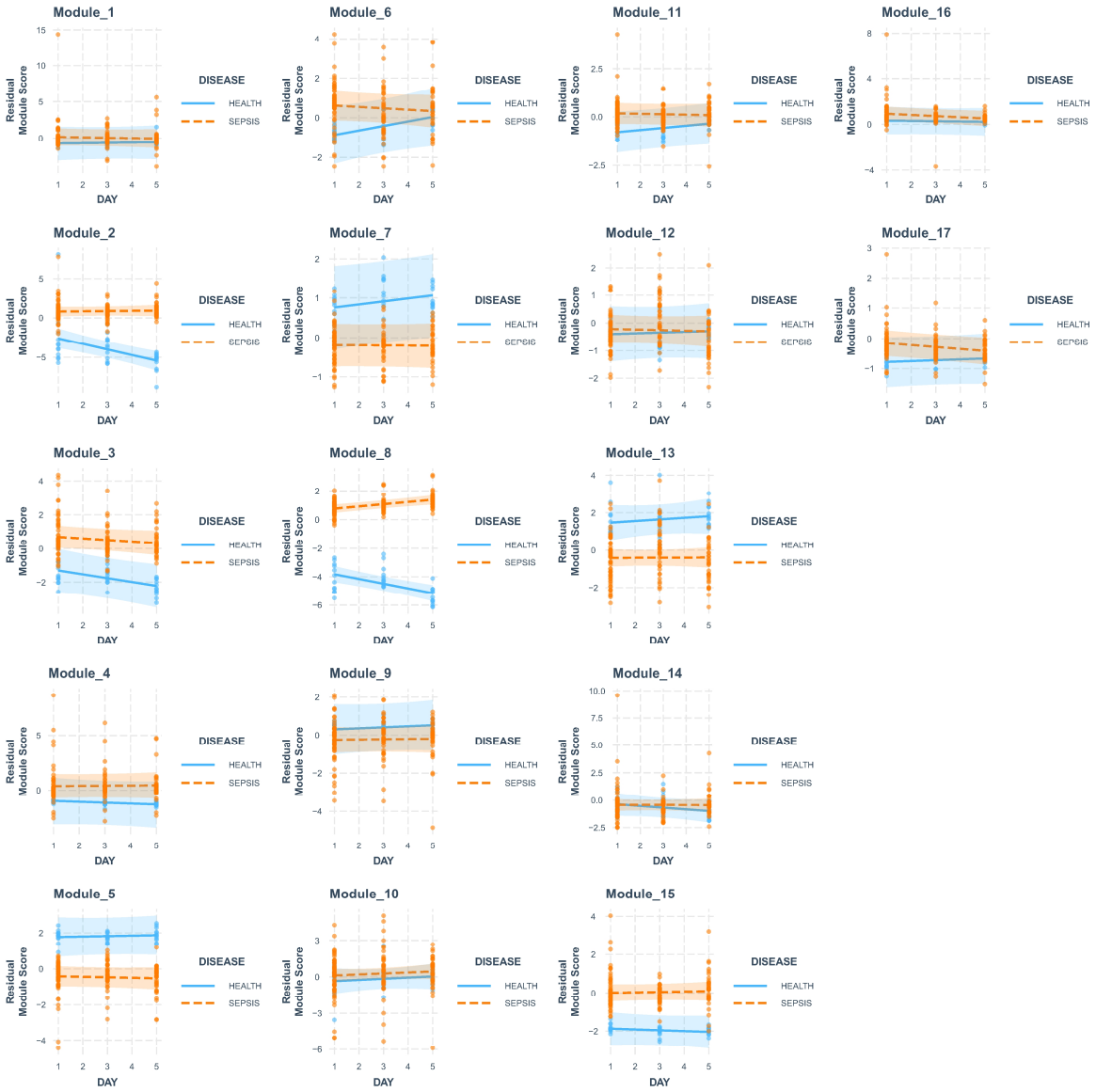


Figure 92. Partial residual plots showing the change in TCRAB module score in sepsis compared to health, across time, while controlling for covariates.

A partial residual plot showing the interaction between time point and disease status and their relationship with TCRAB module score, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (D*TASC model). 95% confidence intervals are shown.

5.3.4.2.2 TCRGD

One third of TCRGD repertoire modules (4, 5 and 6) exhibited a significant difference between health and sepsis ($p<0.05$) and two additional modules (2 and 7) had a significant disease by time interaction [**Figure 93**, **Figure 94**, **Table D. 2-Table D. 4**]. All five of

these modules explained significantly more variance than both the null and covariate only model ($p<0.05$). In the case of module 2 and 7, the effect of time was significant in healthy controls, which can again be attributed to the substantial variation in read depth among healthy controls libraries, particularly for TCRD library 27 [**Figure 81**]. However, for module 7, there was still a significant difference in module score between health and sepsis at time points 3 and 5 suggesting a genuine biological difference [**Figure 93.B**]. Unlike observed for the BCR/TCRAB repertoire, there was no evidence of association with either age, sex or Charlson Index. However, there were a smaller number of matched TCRG and TCRD samples (n=158), which may have limited statistical power to detect these associations.

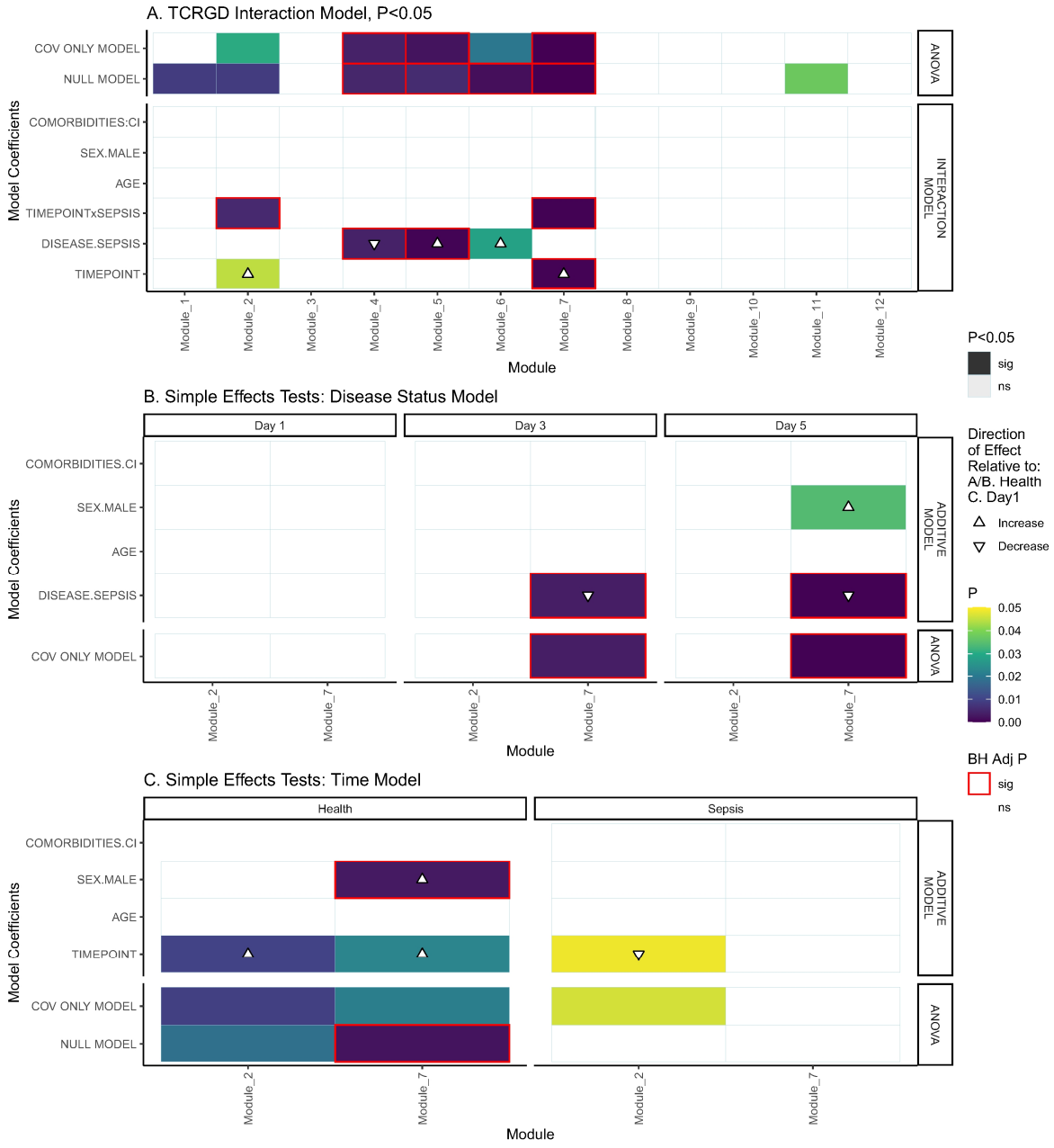


Figure 93. Linear mixed effect models investigating the effect of time and disease status on TCRGD module score.

A) The D*TASC model was applied to each module separately and significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the interaction model explained significantly more variance than either a null model or covariate only model. **B-C)**. Simple effects tests were conducted on modules with a significant interaction between time point and disease status and results are shown for: **B)** the effect of disease status (health or sepsis) at days 1, 3 and 5 and **C)** the effect of time on module score for either health or disease.

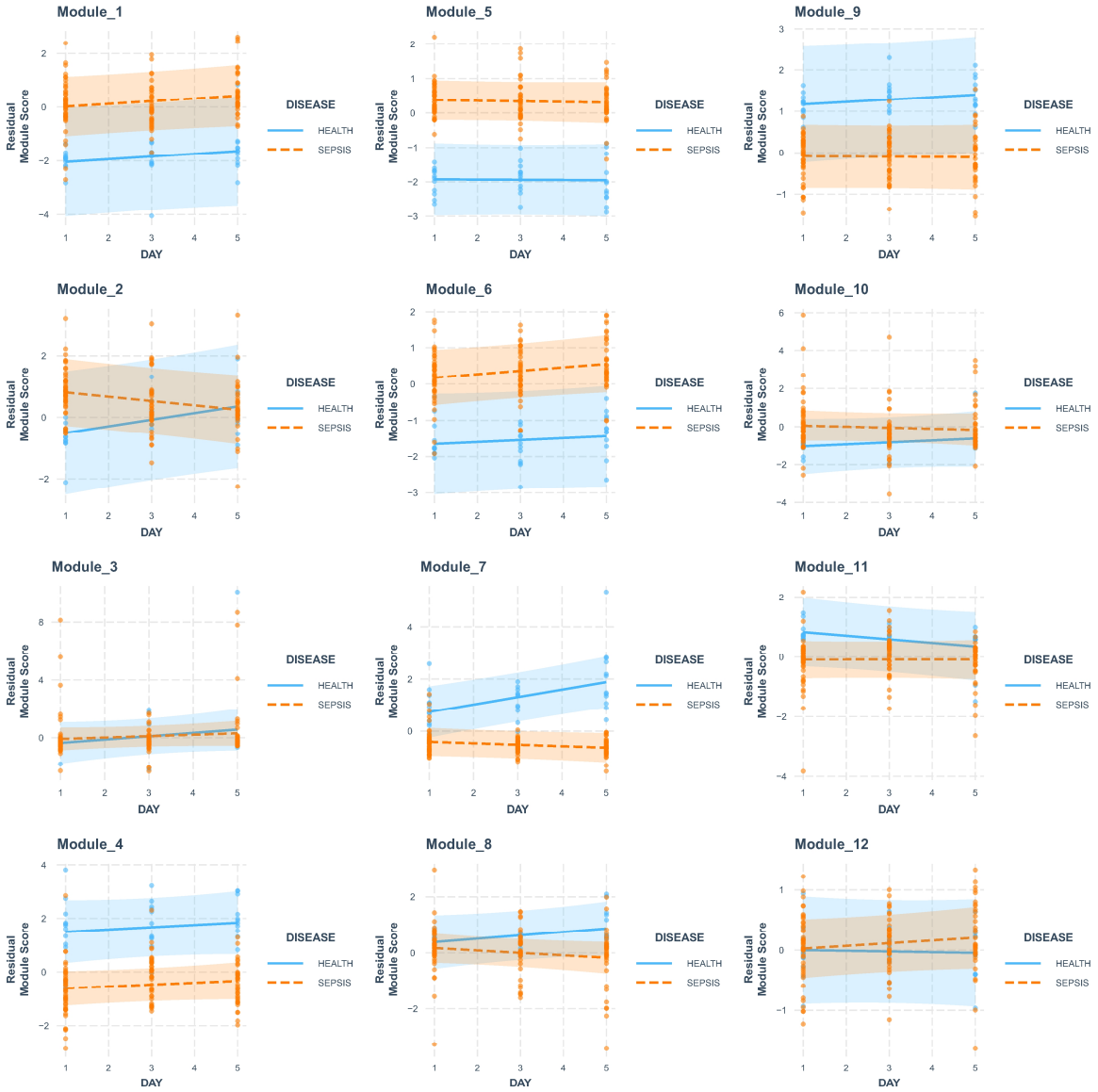


Figure 94. Partial residual plots showing the change in TCRGD module score in sepsis compared to health, across time, while controlling for covariates.

A partial residual plot showing the interaction between time point and disease status and their relationship with TCRGD module score, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (D*TASC model). 95% confidence intervals are shown.

5.3.4.3 Sepsis Mortality

To detect significant associations between TCR module score and mortality, interaction analysis was conducted using the M*TASC model, adjusting for covariates. Sepsis patients

were stratified according to 7-day and 6-month survival status. In the case of a significant interaction term, follow-up simple effects tests were performed.

5.3.4.3.1 TCRAB

There were 7 TCRAB modules that showed significant difference between survivors and 7-day non-survivors (2, 7, 8, 9, 14, 16 and 17) [**Figure 95**, **Figure 96**, **Table D. 5-Table D. 7**]. Of these, module 7 also showed a significant difference between survivors and late-death non-survivors, but in the opposite direction. However, the variance explained by the M*ATSC model for module 2 did not explain significantly more variance than the covariate-only model, indicating that mortality explains a relatively small proportion of variation in the data.

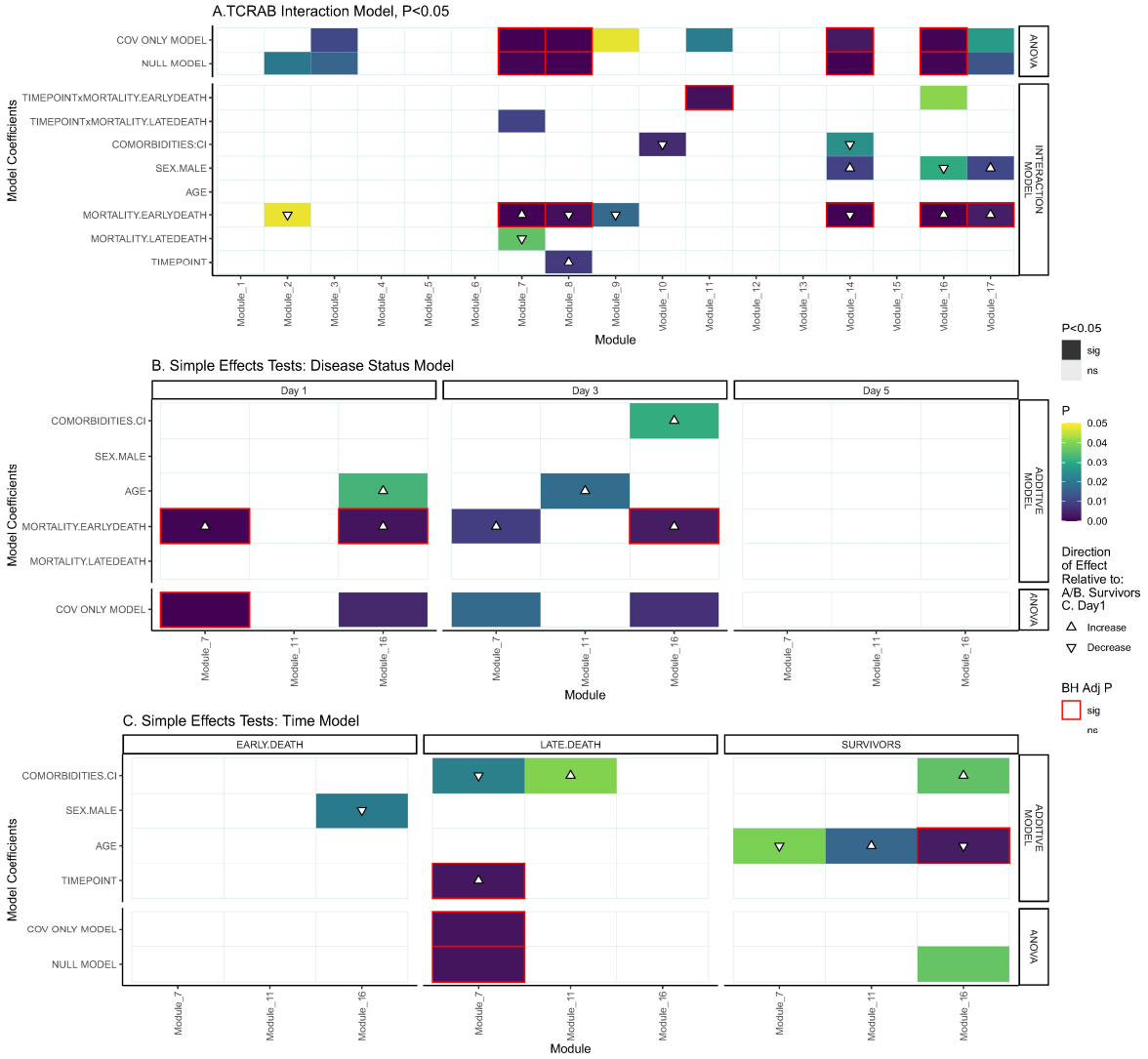


Figure 95. Linear mixed effect models investigating the effect of time and sepsis mortality group on TCRAB module score.

A) The M*TASC model was applied to each module separately and significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the interaction model explained significantly more variance than either a null model or covariate only model. Simple effects tests were conducted on modules with a significant interaction between time point and mortality; results are shown for: **B)** the effect of mortality at days 1, 3 and 5 and **C)** the effect of time on module score for each mortality group.

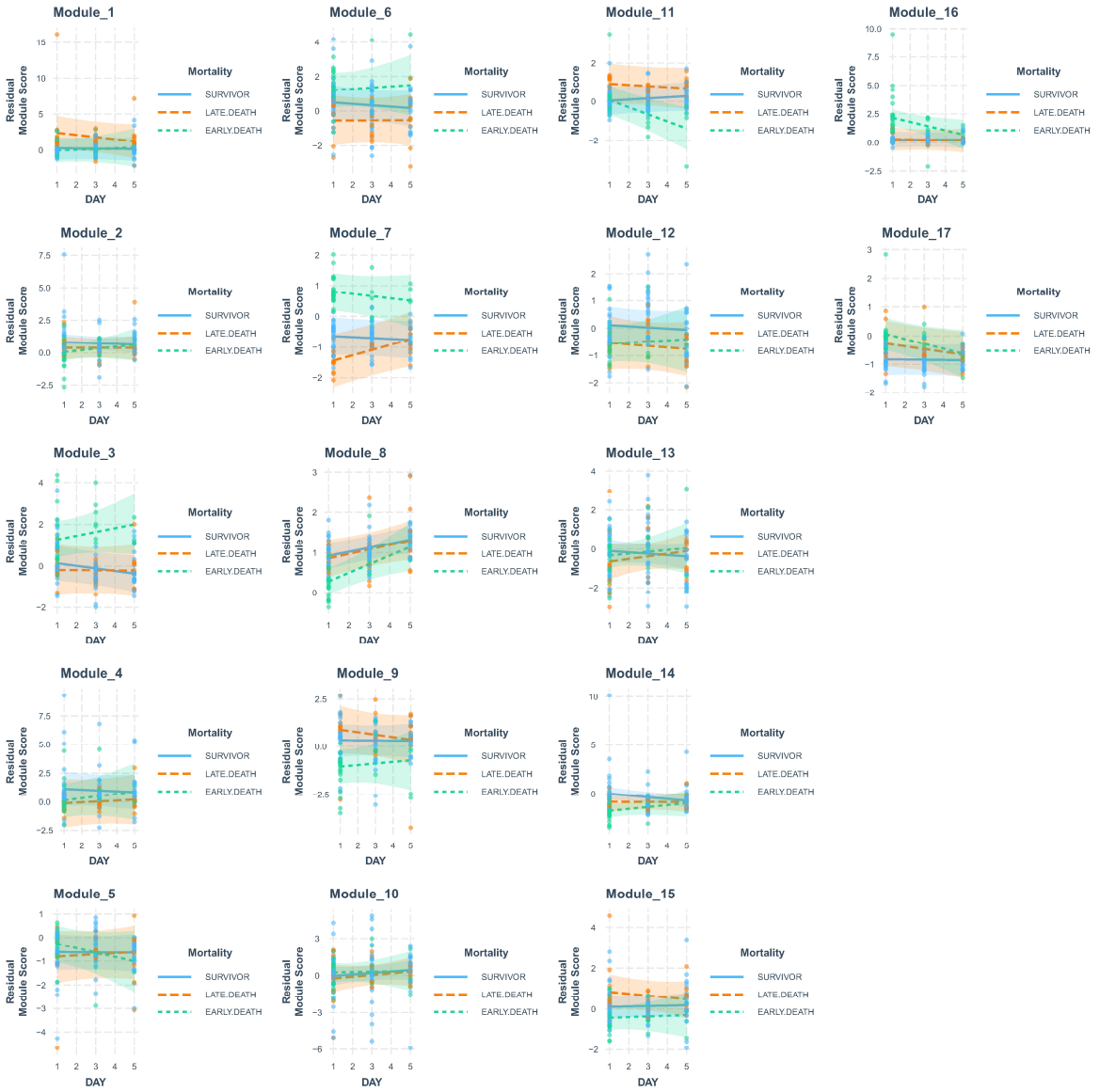


Figure 96. Partial residual plots showing the changes in TCRAB module score over time, in sepsis mortality groups, while controlling for covariates.

A partial residual plot showing the interaction between time point and sepsis mortality group and their relationship with TCRAB module score, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (M*TASC model). 95% confidence intervals are shown.

5.3.4.3.2 TCRGD

Five TCRGD modules (1, 5, 7, 8 and 12) displayed a significant difference in module score between sepsis survivors and 7-day non-survivors [**Figure 97**, **Figure 98**, **Table D. 5-Table D. 7**]. However, module 1 did not explain significantly more variance than either

the null model or the covariate only module. Module 7 had a significant interaction term, however when analysed separately using simple effects test, module 7 score was only significantly different between survivors and 7-day non-survivors at day 1, and no difference was seen between survivors and late–death non-survivors. The lack of statistical significance may be the result of sample drop-out due to death.

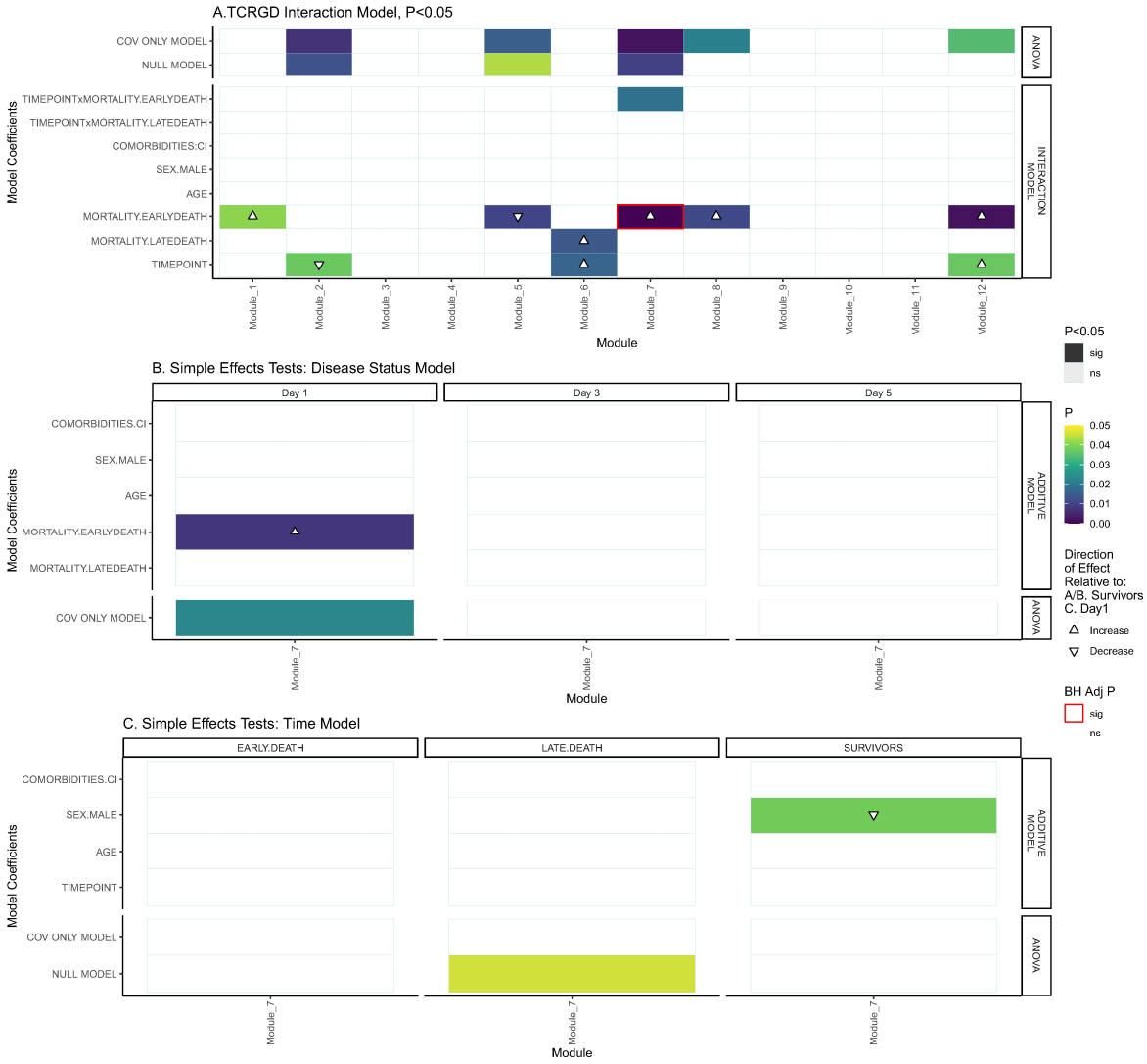


Figure 97. Linear mixed effect models investigating the effect of time and sepsis mortality group on TCRGD module score.

A) The M*TASC model was applied to each module separately and significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the interaction model explained significantly more variance than either a null model or covariate only model. Simple effects tests were conducted on modules with a significant interaction between time point and mortality; results are shown for: **B)** the effect of mortality at days 1, 3 and 5 and **C)** the effect of time on module score for each mortality group.

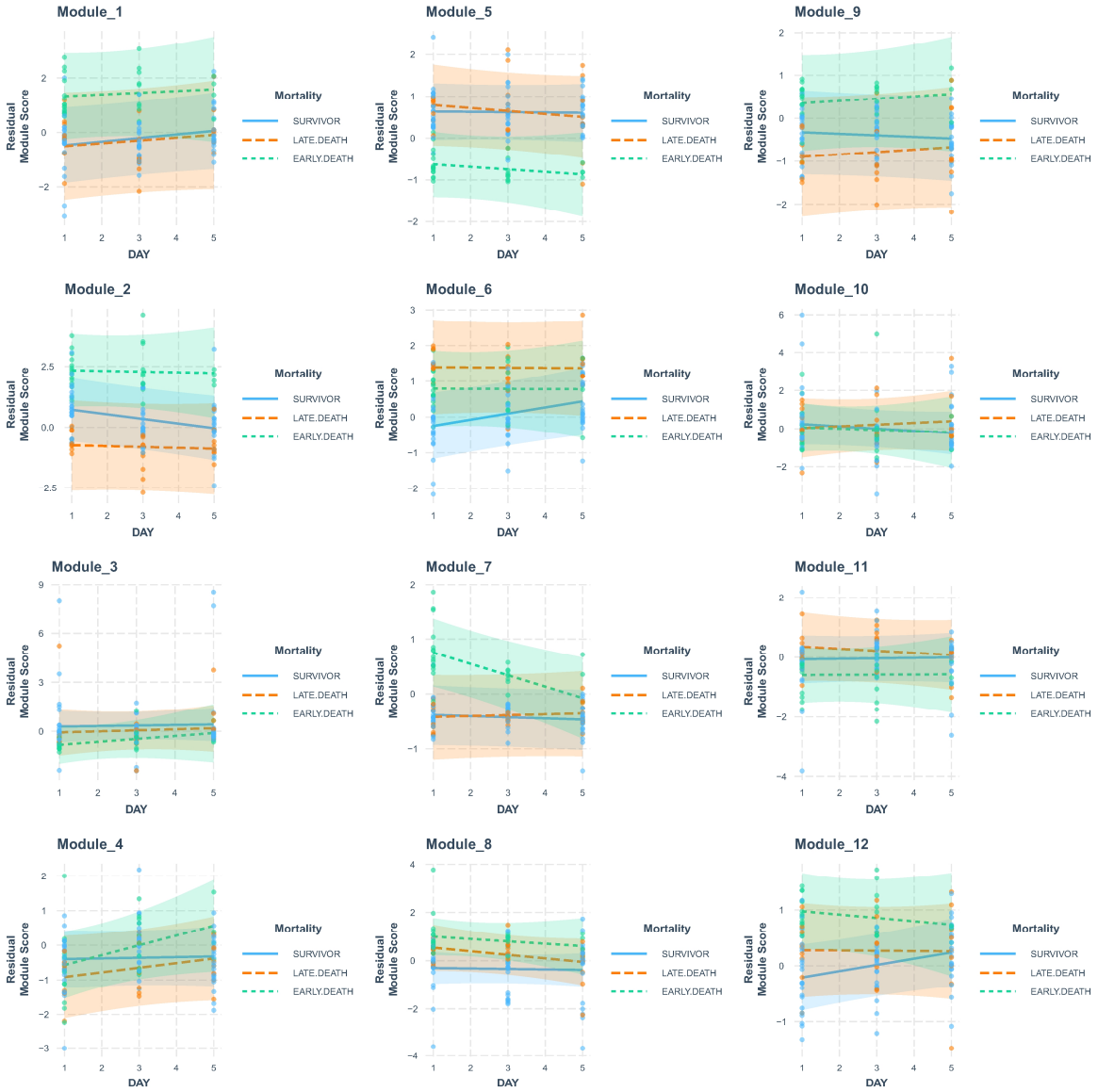


Figure 98. Partial residual plots showing the changes in TCRGD module score over time, in sepsis mortality groups, while controlling for covariates.

A partial residual plot showing the interaction between time point and sepsis mortality group and their relationship with TCRGD module score, controlling for the estimated effects of the clinical covariates age, sex and Charlson Index and the random effect term for non-independence of samples (M*TASC model). 95% confidence intervals are shown.

5.3.4.4 Inter-Module Associations

RMCORR was used to investigate intra-receptor module relationships, accounting for non-independence of samples (Bakdash and Marusich, 2017).

5.3.4.4.1 TCRAB

Visualisation of the TCRAB module correlation network, reveals that modules are relatively isolated, suggesting that cluster resolution was appropriate and there is minimal overlap between module biology. A single cluster of positively correlated modules (2, 3, 8, and 14), all related to V gene usage, was observed which may suggest paired expansion of specific TRAV and TRBV genes [**Figure 99**]. The strong negative association between module 5 and module 4 can be explained by the use of inverse clonal expansion measures.

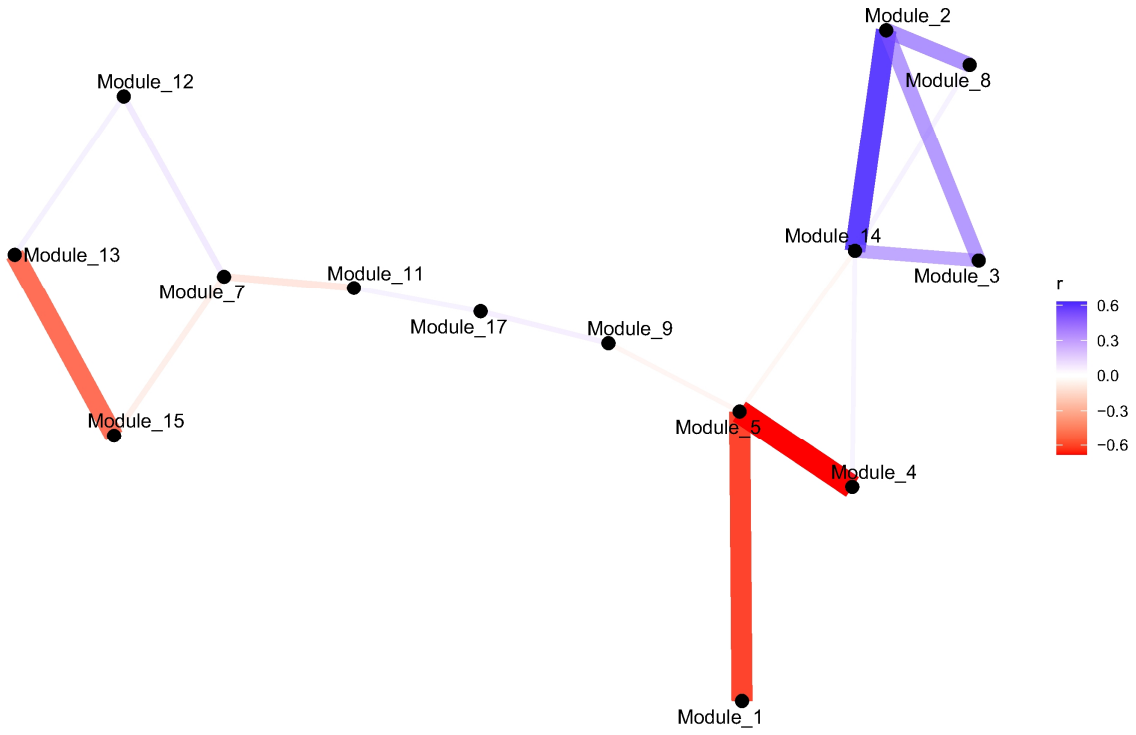


Figure 99. TCRAB module interaction network.

All pairwise correlations between rWGCNA TCRAB modules were calculated using RMCORR to understand high-level organisation. Raw p values were corrected for multiple testing using the Benjamini-Hochberg method. Only inter-module correlations with $R_{rm}>0.3$ and $p_{adj}<0.05$ are shown.

5.3.4.4.2 TCRGD

TCRGD modules cluster into two interconnected networks indicative of distinct biological processes [**Figure 100**]. The separation between the two subnetworks is primarily driven

by whether modules contain descriptive measures (*e.g.* clonal expansion – modules 1, 2, 3, 11 and 10) or quantified V gene usage and the CDR3 region (module 4, 5, 6, 9 and 12). There was a strong positive correlation between module 1 and module 10, which both capture aspects of TCRD clonal expansion. Both of these modules were negatively correlated with the inverse Rényi clonal expansion measures (module 11). Overall, there were few connections between modules (a maximum of 3) suggesting cluster resolution was appropriate.

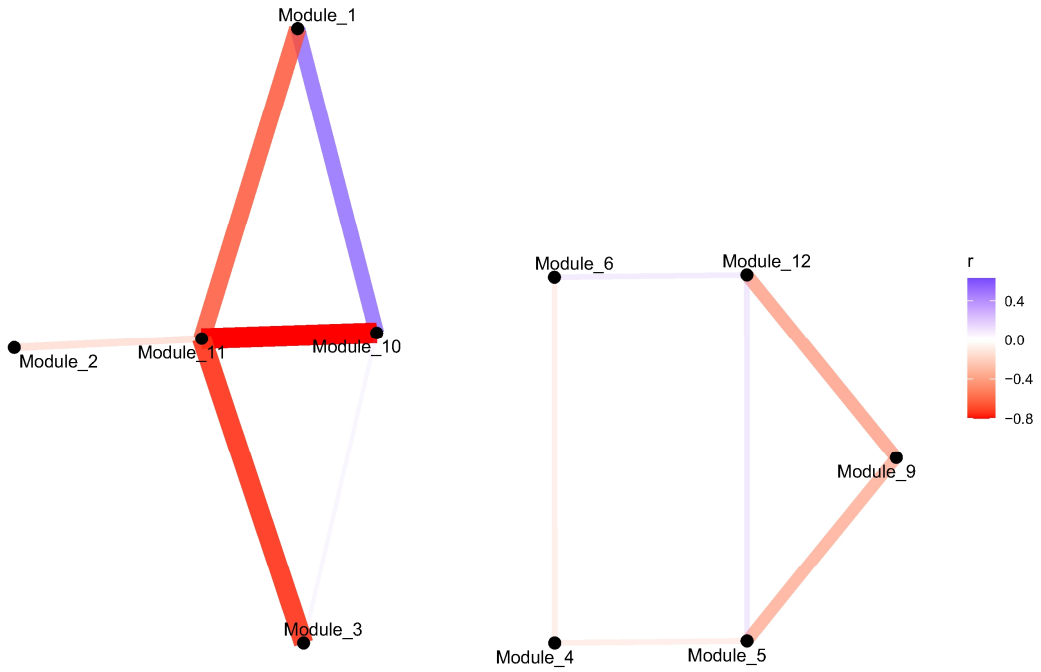


Figure 100. TCRGD module interaction network.

All pairwise correlations between rWGCNA TCRGD modules were calculated using RMCORR to understand high-level organisation. Raw p values were corrected for multiple testing using the Benjamini-Hochberg method. Only inter-module correlations with $R_{rm}>0.3$ and $p_{adj}<0.05$ are shown.

5.3.4.5 Association with Clinical Variables

5.3.4.5.1 TCRAB

The associations between TCRAB module score and clinical variables (either longitudinal or measured once per ICU admission) were analysed using the same approach as for the

BCR repertoire [**Figure 101, Table D. 8, Table D. 9**]. White cell count (WCC), specifically polymorphonuclear (neutrophils, eosinophils, basophils and mast cells) and mononuclear (macrophages and lymphocytes) leukocytes had a significant negative effect on module 9 score. Module 14 was the only module associated with disease severity, when measured using the day 1 SAPS and APACHE II score [**Figure 101.A**]. However, the significant association did not persist when considering the longitudinal SOFA score as a measure of disease severity [**Figure 101.B**]. Interestingly, module 14 was also associated with detection of *Haemophilus influenzae,* whereas modules 17 and 7 were weakly associated with *Streptococcus pneumoniae* infection. This suggests the possibility of distinct TCRαβ repertoire responses between different bacterial causes of sepsis, which provides a promising foundation for further analysis.

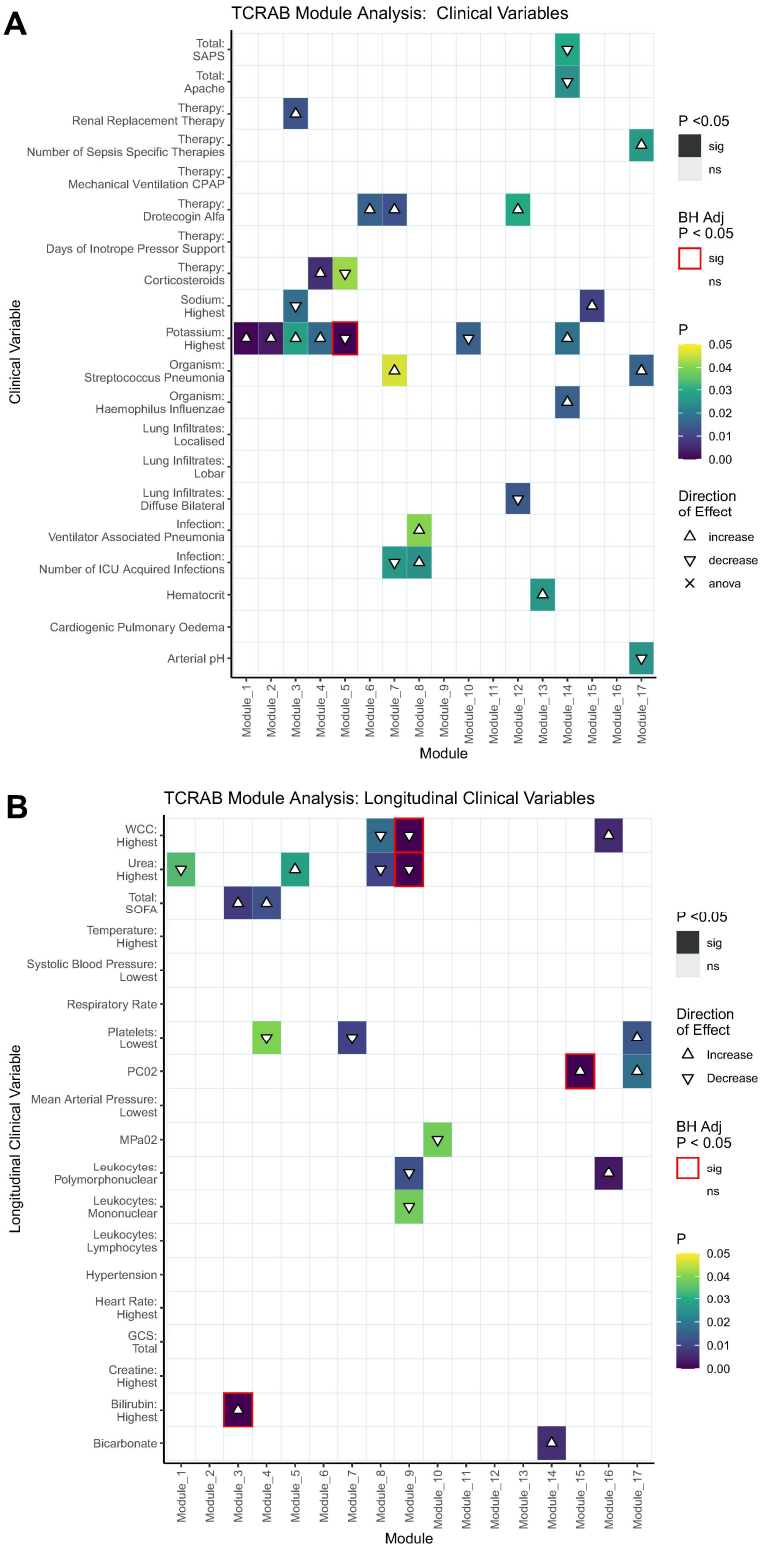


Figure 101. The effect of A) clinical variables measured once per ICU admission on day 1 TCRAB module score and B) longitudinal clinical variables on TCRAB module score (all time points).

For binary clinical variables *e.g.* yes or no for presence of lung infiltrates, the direction of effect is reported relative to those without. Significant model effect terms ($p<0.05$) are shown. The BH multiple testing correction was applied to all calculated p values; significant terms after correction are outlined in red.

5.3.4.5.2 TCRGD

There were several significant associations between TCRGD module scores and clinical variables recorded in the eCRF [**Figure 102**, **Table D. 8, Table D. 9**]. Module 10, which represents TCRD clonal expansion, was positively associated with the number of ICU-acquired infections, ventilator-associated pneumonia, and *Streptococcus pneumoniae* infection. Additionally, across all three time points, white cell count (WCC) and systolic blood pressure exhibited a negative effect on module 10, making it an intriguing module for further investigation. Unique to the TCRγδ repertoire, there were five modules (approximately 40%) that were significantly associated with *Streptococcus pneumoniae* infection. Previous research has shown that γδ T cells play a role in regulating macrophages and dendritic cells during the resolution of *Streptococcus pneumonia*e mediated lung inflammation and have been observed to increase in the lungs of infected mice without a corresponding increase in lymphoid tissue (Kirby et al., 2007, Nakasone et al., 2007, Ivanov et al., 2014).



Figure 102. The effect of A) clinical variables measured once per ICU admission on day 1 TCRGD module score and B) longitudinal clinical variables on TCRGD module score (all time points).

For binary clinical variables *e.g.* yes or no for presence of lung infiltrates, the direction of effect is reported relative to those without. Significant model effect terms ($p<0.05$) are shown. The BH multiple testing correction was applied to all calculated p values; significant terms after correction are outlined in red.

5.3.4.6 Integrative Analysis

Post-QC there was a total of 169 samples with matched TCRAB repertoire data and bulk RNA sequencing gene expression data, and 154 samples with matched TCRGD repertoire data and bulk RNA sequencing gene expression data. These samples were used in integrative analysis, which will be described in the following sections.

5.3.4.6.1 Association with Sepsis Response Signatures:

5.3.4.6.1.1 TCRAB

SRSq score had a significant negative effect on TCRAB module 9, having analysed sepsis samples using the TSqASC linear mixed effects model ($p<0.05$) [**Figure 103**, **Table D. 10**]. This can be interpreted as increased immunosuppression is associated with a decrease in module 9 score. Interestingly, there was minimal correlation between module 9 score and the other TCRAB modules [**Figure 99**]. Therefore module 9 captures a distinct aspect of TCRαβ repertoire biology. However, the proportion of variance explained by this model was not significantly greater than the null model, meaning SRSq score explains only a small portion of the total variance in module 9. Module 9 is comprised of features that describe the CDR3 region of the TCR α and β chains, including CDR3 charge and length [**Table 29**].

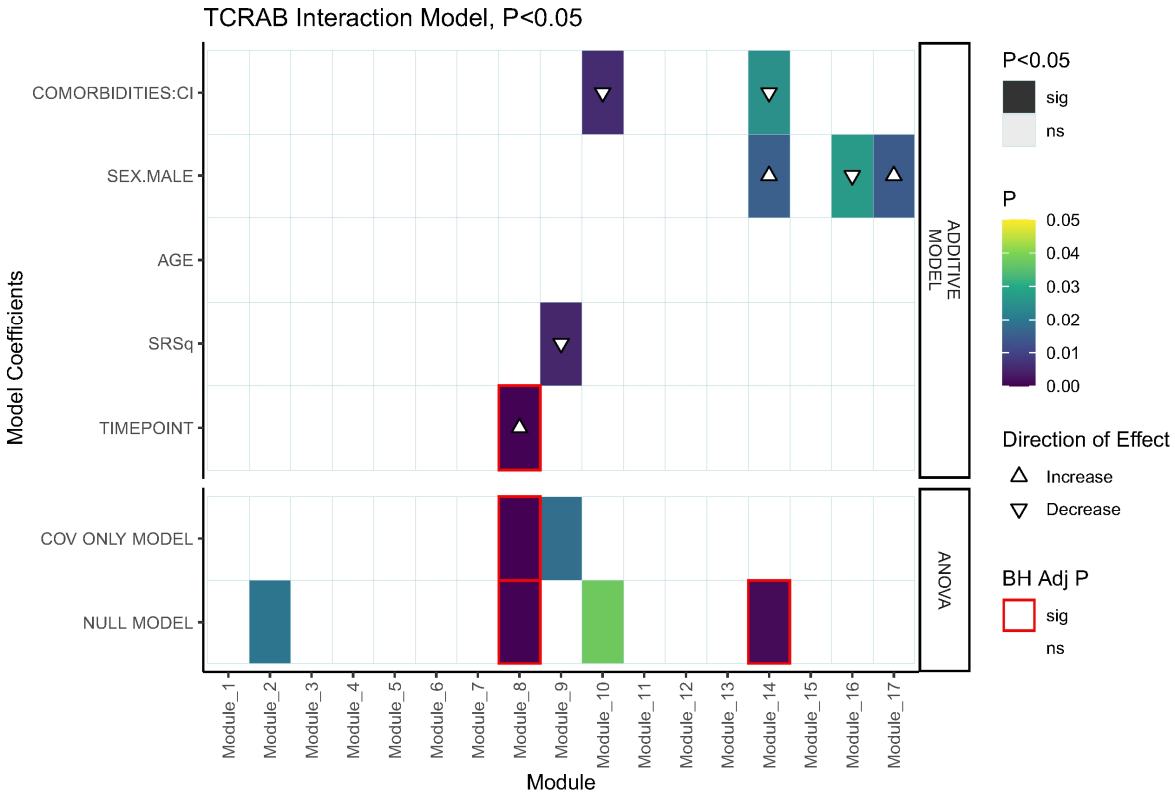


Figure 103. Linear mixed effects models exploring the effect of SRSq score on TCRAB module score (TSqASC).

Significant model effect terms ($p<0.05$) are shown. The Benjamini-Hochberg multiple-testing correction was applied to all calculated p values; significant terms after correction are outlined in red. ANOVA was used to test whether the model performed significantly better than either a null model or covariate only model under the parsimony principle.

5.3.4.6.1.2 TCRGD

There were no TCRGD modules associated with sepsis response signature quantitative score ($p<0.05$). This is not surprising considering TCRGD cells comprise just 5% of the total T cell population, therefore their contribution to a whole blood derived transcriptomic score will be minimal [**Table D. 10**].

5.3.4.6.2 Association with CIBERSORTx Estimated Cell Fractions

To examine the relationship between TCR module scores and CIBERSORTx estimated cell fractions, I employed quantitative trait analysis as performed for the BCR repertoire (Seo

et al., 2016). Associations were considered significant if they met the criteria of an effect size (ES) greater or less than the mean ± 1 standard deviation and a p value less than 0.05.

5.3.4.6.2.1 TCRAB

Reassuringly, multiple associations between different T cell populations and TCRAB repertoire modules were detected [**Figure 104**, **Figure D. 12**]. There were few associations with B cell and plasmablast populations in comparison to the BCR repertoire. The significant association seen between $CD8^+$ T cell population size and module 4, which captures TCRB clonal expansion measures, is evidence of the validity of this analytical approach.

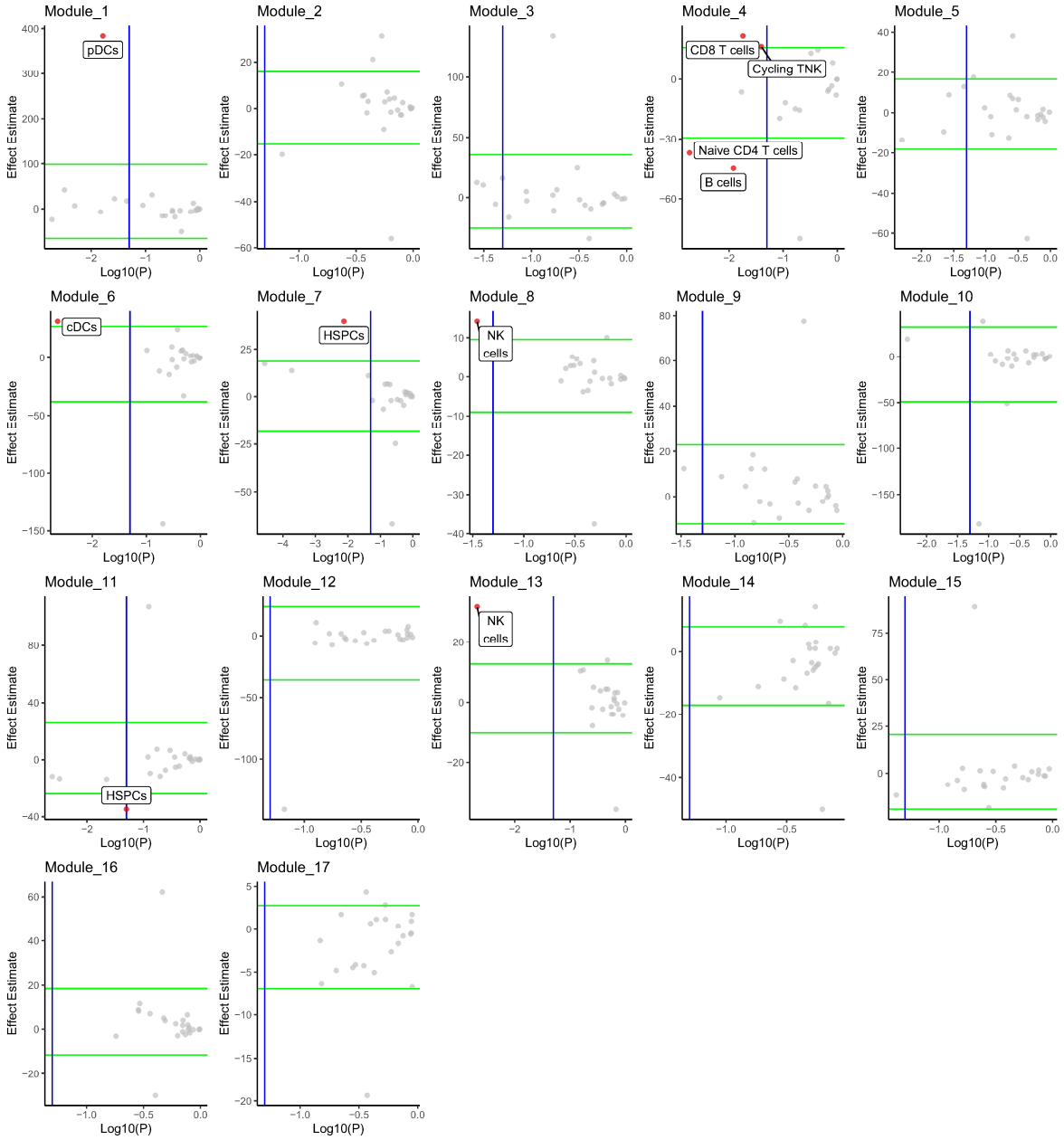


Figure 104. Association between CIBERSORTx estimated cell fractions and TCRAB module score.

Cell fraction was used as the independent variable and module score was the dependent variable. Estimated effect sizes on module score for each cell fraction and their corresponding p values are plotted. Log10(p=0.05) is shown in blue and the mean ± 1 SD is shown in green, cell types which meet these requirements are annotated.

5.3.4.6.2.2 TCRGD

Due to their rarity, γδ T cells were not among the cell types included in the single cell sepsis reference supplied to CIBERSORTx. This may explain the limited number of associations between cell fractions and TCRGD module score (independent variable). However, there

were several interesting relationships, where TCRGD score had a significant effect on the population size of other immune cell types. For example, an increase in module 2 (TRG clonality) score was associated with a decrease in the estimated number of B cells and naïve $CD8^+$ T cells [**Figure 105**, **Figure D. 13**].

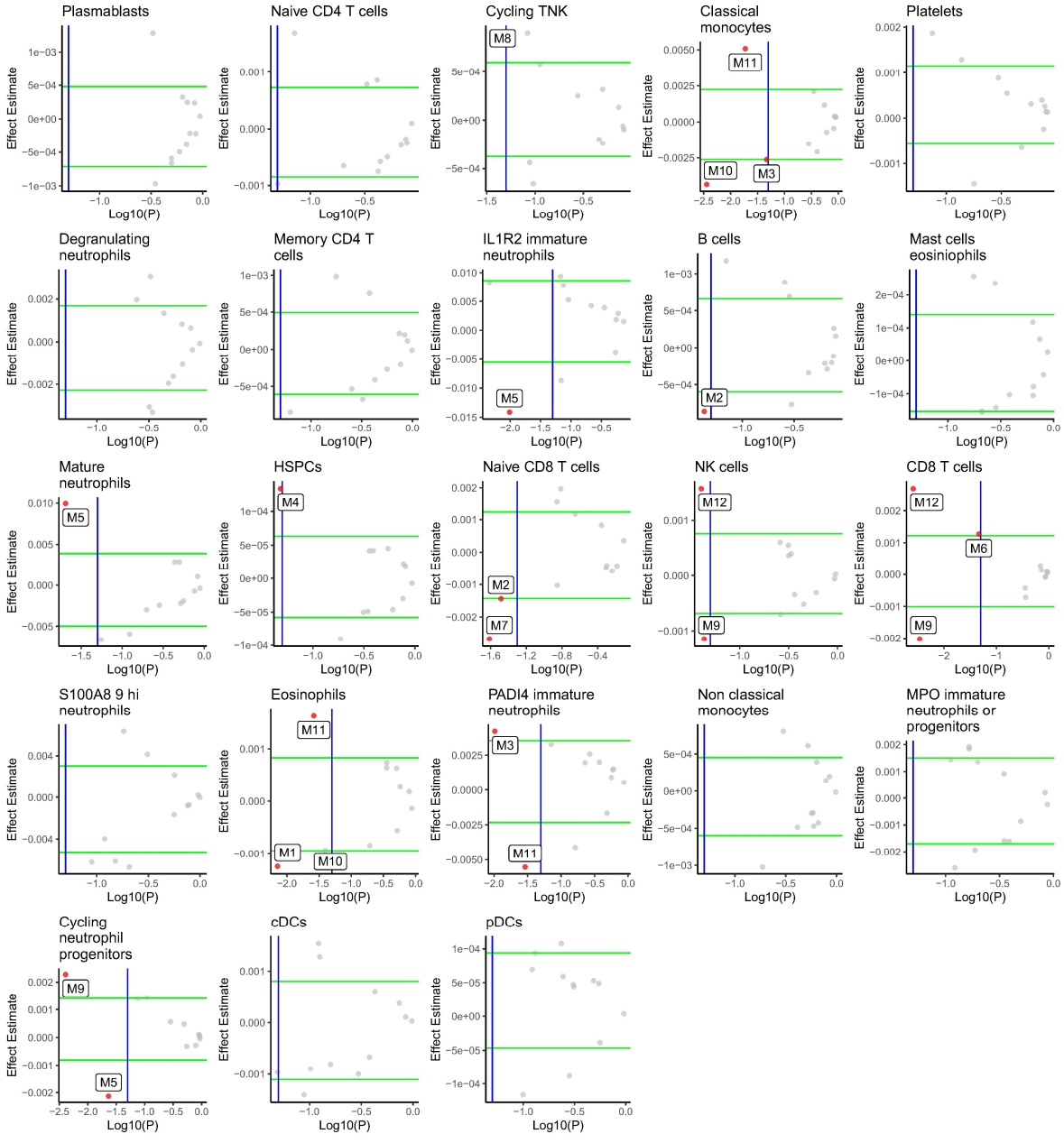


Figure 105. Association between TCRGD module score and CIBERSORTx estimated cell fractions.

Module score was used as the independent variable and cell fraction was the dependent variable. Estimated effect sizes on cell fraction size for each module and their corresponding p values are plotted. Log10(p=0.05) is shown in blue and the mean ± 1 SD is shown in green, cell types which meet these requirements are annotated.

5.3.4.7 Principal Component Analysis

5.3.4.7.1 TCRAB

PCA was applied to the imputed scaled TCRAB matrix and visualisation reveals clear separation between healthy controls and sepsis patients, with limited overlap in PC1 coordinates [**Figure 106.A**]. Separation between health and sepsis in PC1 and PC2 is largely driven by TRBV gene usage and TCRA clonality, as discussed in sections in **5.3.5.1** and **5.3.5.2** [**Figure 107**].

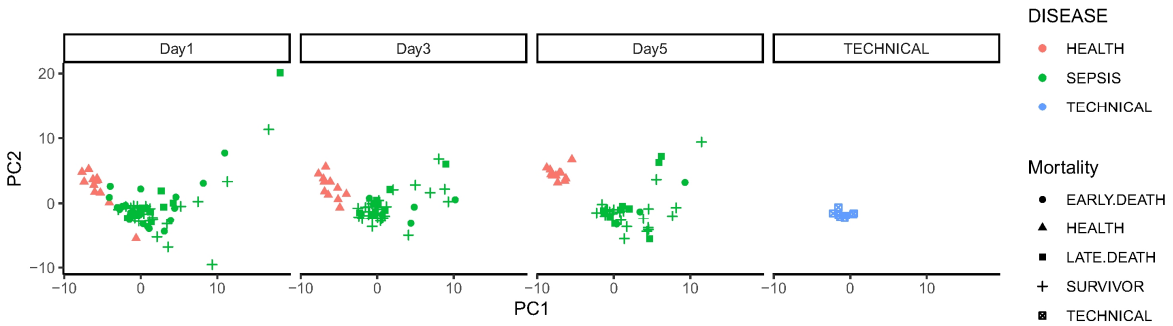


Figure 106. TCRAB principal component analysis.

PCA was applied to the scaled imputed TCRAB repertoire feature matrix directly. The first two components are plotted and points are coloured according to disease status. The shape corresponds to the sepsis mortality group. Clear separation of healthy volunteer and sepsis samples is observed.

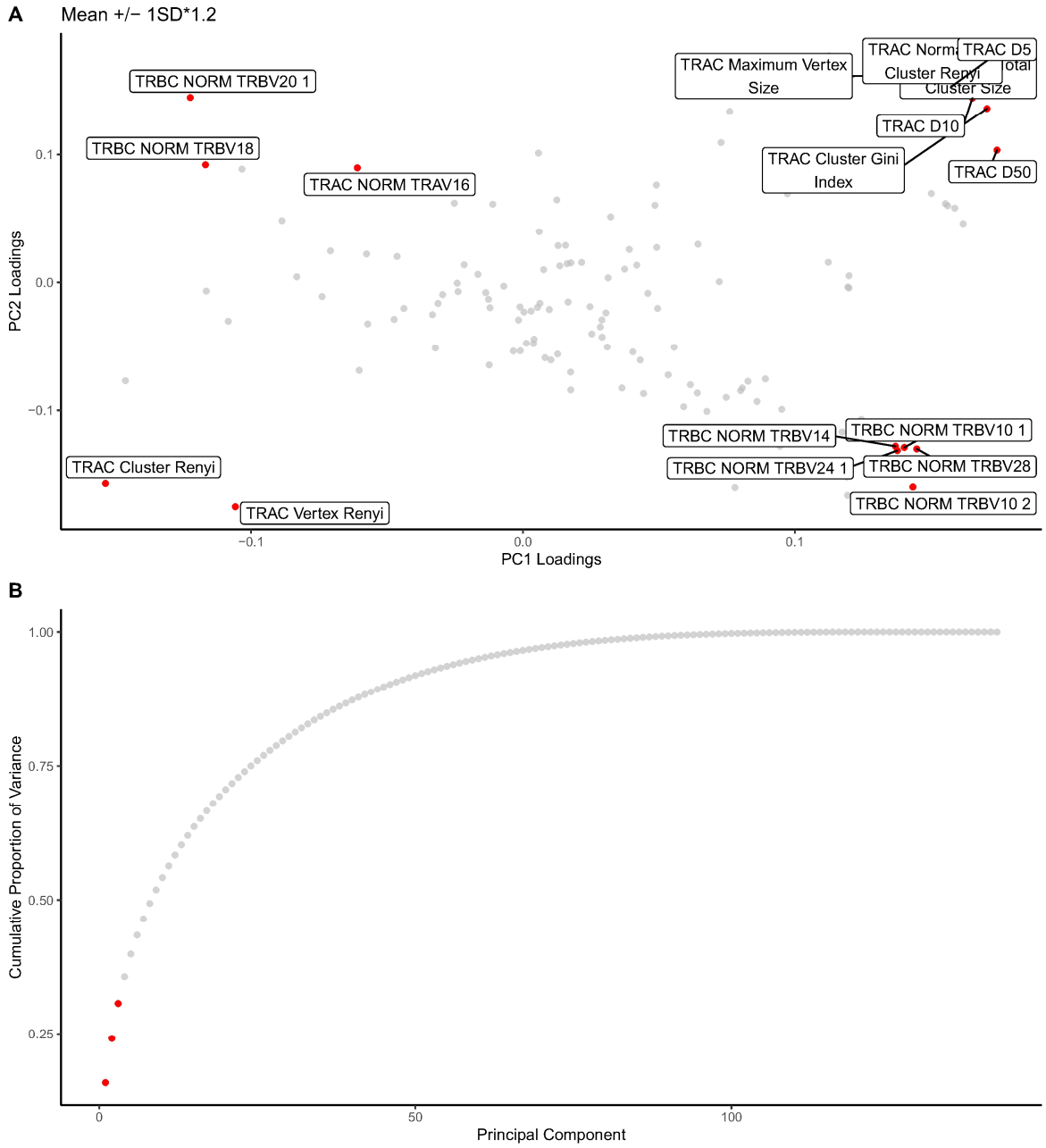


Figure 107. Separation between health and disease is driven by a subset of TCRAB repertoire features.

PCA was applied to the scaled imputed TCRAB repertoire feature matrix directly. **A)** PC1 and PC2 feature loading scores were plotted to identify the subset of features which drive separation of health and sepsis in **Figure 106.** Features are highlighted if their loading score is outside of the mean ± 1 SD x1.2 for both components. **B)** Cumulative proportion of variance explained by each PC. Points are highlighted if each PC explains more than 5% of variance in the data.

5.3.4.7.2 TCRGD

Similarly, PCA was applied to the scaled imputed TCRGD matrix. Good separation between sepsis and health was observed, although the technical samples do not cluster as tightly. This is most likely because there are fewer TRGV and TRDV genes therefore the repertoire is less 'personalised'. PC1 and PC2 explain almost 40% of variance in the data, and separation was largely driven by TCRG clonal expansion measures and differences between TRD1 and TRD2 usage [**Figure 108**, **Figure 109**]. Interestingly, TRG clonal expansion measures did not differ significantly between health and disease when modelled as part of TRGD module analysis, suggesting the dataset was underpowered to detect small effect sizes at statistical significance. Collectively however, these features discriminate between health and disease [**Figure 93**].

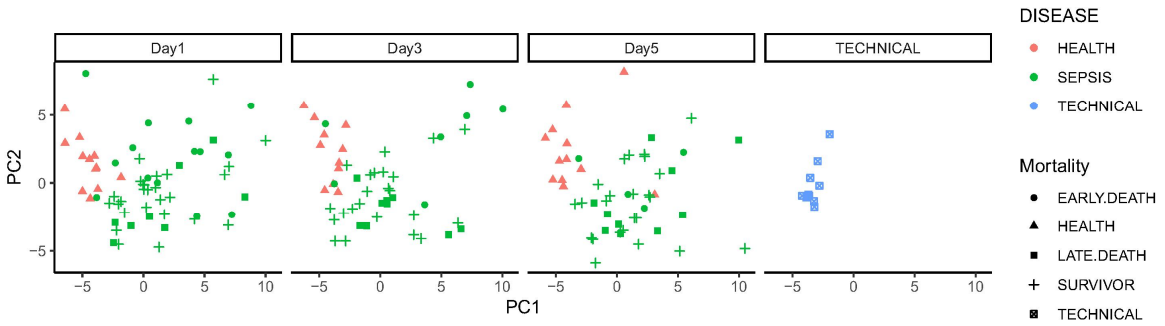


Figure 108. TCRGD principal component analysis.

PCA was applied to the scaled imputed TCRGD repertoire feature matrix directly. The first two components are plotted and points are coloured according to disease status. The shape corresponds to the sepsis mortality group. Clear separation of healthy volunteer and sepsis samples is observed.

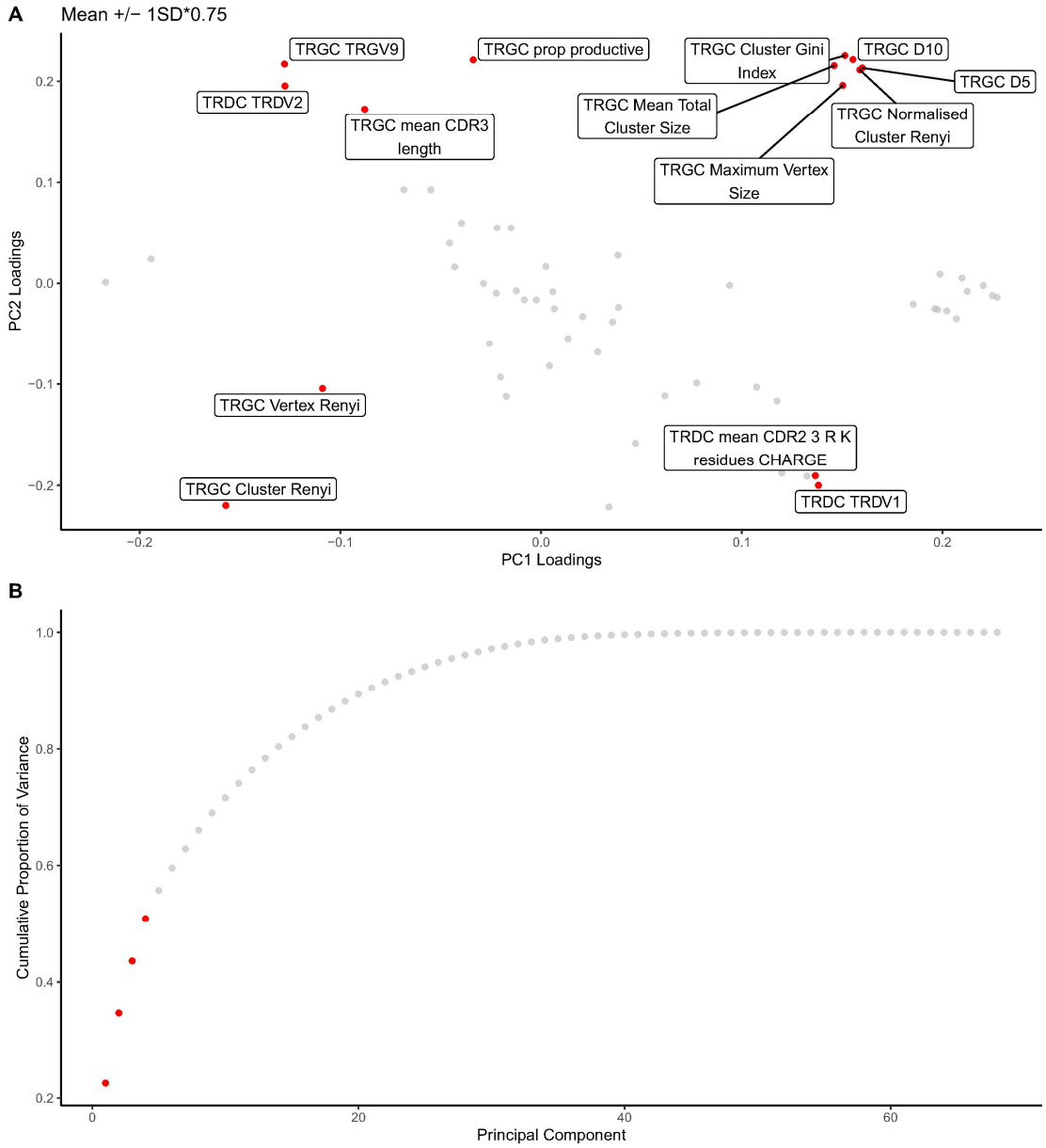


Figure 109. Separation between health and disease is driven by a subset of TCRGD repertoire features.

PCA was applied to the scaled imputed TCRGD repertoire feature matrix directly. **A)** PC1 and PC2 feature loading scores were plotted to identify the subset of features which drive separation of health and sepsis in **Figure 108**. Features are highlighted if their loading score is outside of the mean ± 1 SD x0.75 for both components. **B)** Cumulative proportion of variance explained by each PC. Points are highlighted if each PC explains more than 5% of variance in the data.

5.3.5 Characterising TCR Repertoire Modules in Sepsis

5.3.5.1 Changes in TRAV and TRBV Gene Usage Distinguish Health and Sepsis

Of the 6 TCRAB modules where there was a significant difference in module score between health and sepsis, 5 out of 6 modules (2, 3, 8, 13 and 15) captured TRBV and TRAV normalised gene usage [**Table 31**]. These modules included a greater number of TRBV genes than TRAV genes, not surprising given the greater number of functional TRBV genes.

Although there was typically a good correlation between feature scores and the corresponding module eigenvector, the difference between health and sepsis was not always statistically significant for each V gene within each module. As an example, partial residual plots visualising non-significant differential TRBV gene usage between health and disease are shown for module 3 in **Figure 110**. This suggests that changes to TRAV and TRBV gene usage in sepsis are subtle and a larger sample size is needed to achieve statistical significance between health and disease. However, when these small changes are considered collectively using a dimensional reduction approach like rWGCNA, a clear separation between sepsis patients and healthy volunteers emerges.

There were no strong significant correlations between the TRAV gene containing modules (13 and 15) and TRBV gene containing modules (2, 3 and 8) [**Figure 99**]. Therefore, the presence of common TRBV and TRAV pairings across sepsis samples is unlikely, but this would require further validation using paired TCRαβ sequencing.

Module	V Gene	Direction	Module interaction	Module associated genes	CIBERSORTX cell fraction association
2	*TRBV12-4, TRBV10-3, TRBV12-5, TRBV14, TRBV19, TRBV4-2, TRBV5-6*	Increased in sepsis	Positive correlation with module 8 and module 3	***DNASE1L3***	
3	*TRBV11-2, TRBV29-1, TRBV5-4, TRBV7-6*	Increased in sepsis	Positive correlation with module 2	**64 genes. Pathways:** mitotic prometaphase and M Phase	**Degranulating neutrophils**
8	*TRBV10-1, TRBV10-2, TRBV15, TRBV24-1, TRBV28*	Increased in sepsis	Positive correlation with module 2	**None**	**NK Cells**
13	*TRAV8-6, TRAV8-1*	Decreased in sepsis	Positive correlation with module 15	**None**	**NK cells**
15	*TRAV4, TRAV40, TRAV30*	Increased in sepsis	Positive correlation with module 13	**None**	

Table 31. Significant perturbations in TRA and TRB V gene usage in sepsis.

TCRAB modules 2, 3, 8, 13 and 15 differed significantly between health and sepsis and captured predominantly changes in TRAV and TRBV gene usage. Linear mixed effects models were used to identify genes significantly associated with TCRAB module score, using bulk RNA sequencing gene expression data. P values were corrected for multiple testing using the Benjamini-Hochberg method. The number of genes significantly associated with module score ($p_{adj}<0.05$ and effect size greater or less than the mean ± 1 standard deviation) are noted. Significant relationships with CIBERSORTx estimated cell fractions are also recorded, along with significant associations between the listed TCRAB modules.

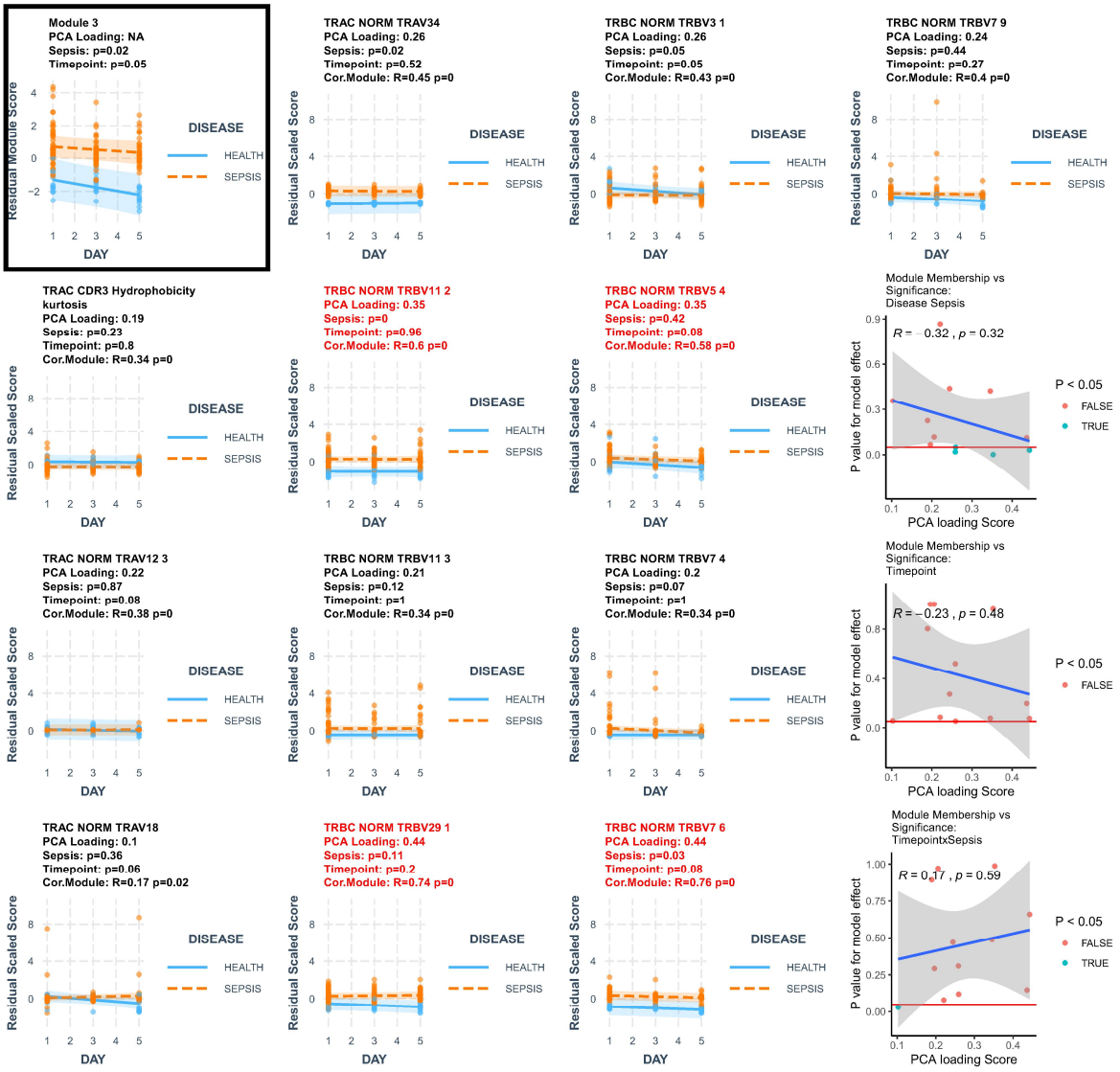


Figure 110. Partial residual plots showing the significant increase in TCRAB module 3 score in sepsis compared to health, and the corresponding changes in TRBV gene usage for the constituent features.

Top left shows the partial residual plot for TCRAB module 3 using the D*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 3 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (sepsis, time point and disease time point interaction). Results are reported to two decimal places.

Quantitative trait analysis identified only a limited number of non-AIR genes with a significant impact on V gene-associated module scores [**Table 31**]. Paired TCRα and β chains are required for the recognition and binding of presented peptide-MHC complexes.

However, TCR signalling is propagated through the associated CD3 complex, comprised of γ/δ/ε/ζ subunits, and relies on immunoreceptor tyrosine-based activation motifs (ITAMs) (Hwang et al., 2020). Therefore, we would not necessarily expect to find a direct association between V gene usage and differential gene expression. However, as the strength of TCR signalling can affect T cell differentiation, if the module captures V genes that have strong affinity for commonly encountered sepsis antigens, we may detect non-AIR genes involved in T cell activation and down-stream differentiation.

Interestingly, there were multiple genes (n=64) significantly associated with module 3 score. Among these genes, two showed particularly strong positive associations: *FOSB* (ES=0.59, p_{adj}=0.00031) and c-Jun (*JUN*) (ES=0.98, p_{adj}=0.00081) [**Figure 111**]. FOSB and c-Jun form the transcription factor dimer AP-1 which is an important regulator of cytokine expression and a modulator of inflammatory disease (Zenz et al., 2008). AP-1 has been shown to bind a recognition site in the 5' untranslated region (UTR) of the CD95 ligand (*CD95L,* ES=0.24, p_{adj}=0.60) gene. Upon TCR/CD3 re-stimulation, CD95L is induced, promoting apoptosis by binding to the CD95 receptor (Fas), which is expressed on various cell types, including chronically activated T cell clones and peripherally activated T and B cells (Iglesias et al., 1999). CD95L triggers apoptosis in CD95L-expressing cells, through the formation of the death-inducing signalling complex and this process contributes to immune homeostasis and self-tolerance (Baumann et al., 2003, Eichhorst et al., 2000). Accordingly, module 3 was associated with a significant decrease in the estimated fraction size of degranulating neutrophils which are highly susceptible to Fas-mediated apoptosis (Liles et al., 1996) [**Figure D. 12**]. The gene *EZH1* had a strong negative association with module 3 score (ES=-1.75, p_{adj}=0.021) and is a component of the polycomb repressive complex 2. It is an important negative regulator of lymphoid commitment during hematopoietic

development, and studies have shown that *EZH1* suppression allows efficient production of mature induced pluripotent stem cell derived T cells with enhanced functionality that display a molecular signature which approximates peripheral blood αβ T cells (Jing et al., 2022).

The detection of non-AIR genes associated with module 3, suggests that the constituent TRBV genes may be biologically significant in sepsis. In agreement, module 3 was the sole module among the five, which had a positive association with disease severity, as measured using the total SOFA score. It was also the only module associated with the administration of renal replacement therapy at day 1 [**Figure 101**]. Module 3 represents TRBV11-2, TRBV29-1, and TRBV5-4 and TRBV7-6 usage. TRBV7-6 recognises two dominant viral epitopes of cytomegalovirus (CMV), which is reactivated in around 30% of sepsis cases (Chen et al., 2017). Hence, this is an interesting avenue for further analysis.

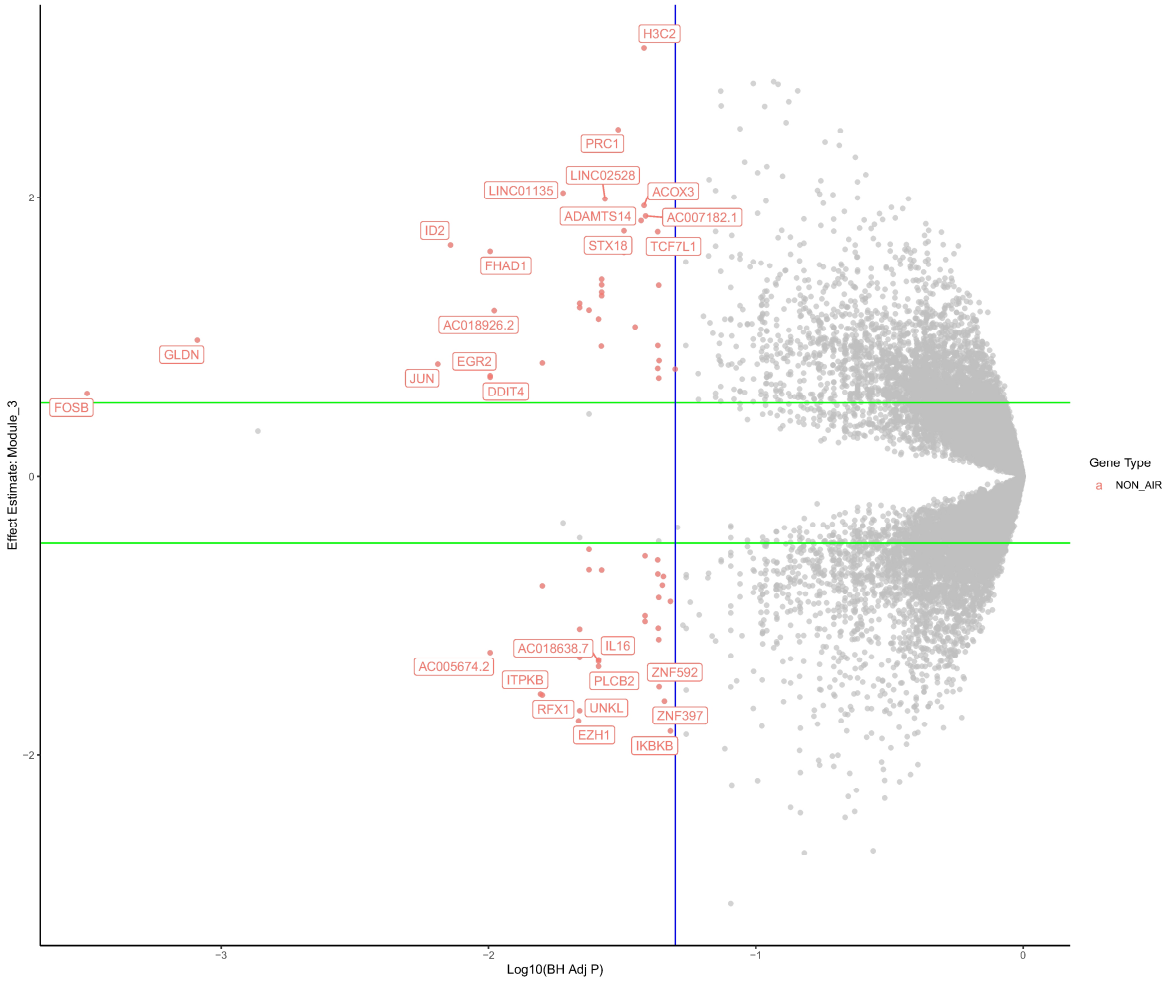


Figure 111. Genes with a significant effect on TCRAB module 3 score, from matched bulk RNA sequencing gene expression data.

Linear mixed effects models were used to identify genes significantly associated with TCRAB module 3 score, using bulk RNA sequencing gene expression data. P values were corrected for multiple testing using the Benjamini-Hochberg method. The log10(p=0.05) is shown by the blue line and the mean effect size ± 1 standard deviation are shown in green. Genes are annotated as either adaptive immune repertoire genes (AIR) or non-AIR genes.

Substantial expansion of TRBV11-2 expressing T cells (up to 24% of clonal T cells) has been reported in children with multisystem inflammatory syndrome (MIS-C) (Porritt et al., 2021, Hoste et al., 2021). MIS-C is associated with SARS-CoV-2 infection and resembles toxic shock syndrome which is triggered by bacterial superantigens. Patients show elevated levels of soluble biomarkers linked to the recruitment and activation of monocytes, neutrophils and septic shock and a monocyte cell MS1 state phenotypically similar to the one in bacterial sepsis (Sacco et al., 2022). Intriguingly, children with TRBV11-2 expansion

are enriched for 3 HLA class I alleles (HLA A*02, B*35, C*04) suggesting genetic susceptibility. TRBV11-2 usage correlated with MIS-C disease severity, as observed by the positive correlation between TCRAB module 3 and sepsis SOFA score and renal replacement in this cohort. Porritt et al. (2021) propose that the interaction between TRBV11-2 and the superantigen-like motif of the SARS-CoV-2 spike glycoprotein is driven by a polyacidic residues and could represent a CDR3-independent expansion. These findings were replicated by Sacco et al. (2022), who also observed that TRBV11-2 expansion was restricted to MIS-C samples collected shortly after hospitalisation, with a rapid decline in TRBV11-2 gene usage after seven days. Moreover, the transcriptional signature of TRBV11-2 $CD4^+$ T cells exhibited increased expression of genes associated with apoptosis and lymphocyte activation.

The patient samples included in this cohort were processed many years prior to the COVID-19 pandemic, excluding prior exposure to COVID-19 as the cause of the observed increase in TRBV11-2 gene usage. Hence, these findings provide further evidence for the involvement of TRBV11-2 in severe respiratory infections characterised by hyper-inflammation. Future research could investigate the dynamics of TRBV11-2 in sepsis at later time points ($>$7 days) and in convalescent samples to ascertain whether TRBV11-2 gene usage declines during disease. Additionally, HLA haplotyping would be of interest to identify whether there is an association between module 3 score and HLA class I alleles in this larger cohort.

5.3.5.2 Increased TCRAB Clonal Expansion in Sepsis

Module 5 was significantly decreased in sepsis compared to health, encompassing V gene usage and TCRAB clonal expansion measures (vertex Rényi), however unlike the results presented in Venet et al. (2013), a return to baseline health in the five-day period after

admission to ICU was not observed [**Figure 112**]. Cluster and Vertex Rényi are inverse measures, therefore module 5 is interpreted as an increase in clonal expansion in sepsis and a decrease in TRAV22 gene usage. Module 5 displayed a strong negative correlation with module 1 and module 4, which represent non-inverse measures of TCRA and TCRB clonal expansion [**Figure 99**].

Small and statistically non-significant differences in clonality were observed between sepsis and health when each constituent feature of module 5 was modelled individually. Similarly, for module 1 and 4, although no significant difference in module score was found between health and sepsis, a minor increase was consistently observed in sepsis across all constituent clonality features, including cluster size, and the proportion of the repertoire explained by the top 5, 10 and 50 clones [**Figure D. 14**, **Figure D. 15**]. These findings collectively provide evidence of increased αβ T cell clonal expansion in sepsis. However, the study was likely underpowered due to cohort heterogeneity and a small sample size.

A decrease in module 5 score (increased clonality) was associated with a decrease in the estimated cell fraction of naïve $CD4^+$ T cells and an increase in the proportion of cycling T or NK (TNK) cells and $CD8^+$ T cells [**Figure D. 12**]. Similarly, an increase in module 4 score was associated with increased $CD8^+$ T cells, cycling TNK cells and a decrease in naïve $CD4^+$ T cells, suggesting a potential overlap in the biological signals captured by these modules [**Figure 104**]. However, no association with leukocyte count was detected, possibly this effect was masked by changes in the population size of other leukocyte subsets.

$CD8^+$ T cells play a crucial role in controlling and resolving viral infections, and selective T cell clonal expansion has been reported in multiple infectious diseases. Following pathogen clearance most expanded T cells undergo apoptosis, although a subset persists as memory

T cells (Callan et al., 1996, Huang et al., 2019). It would be interesting to profile the TCR repertoire of convalescent sepsis patients to determine whether levels return to normal after recovery from sepsis.

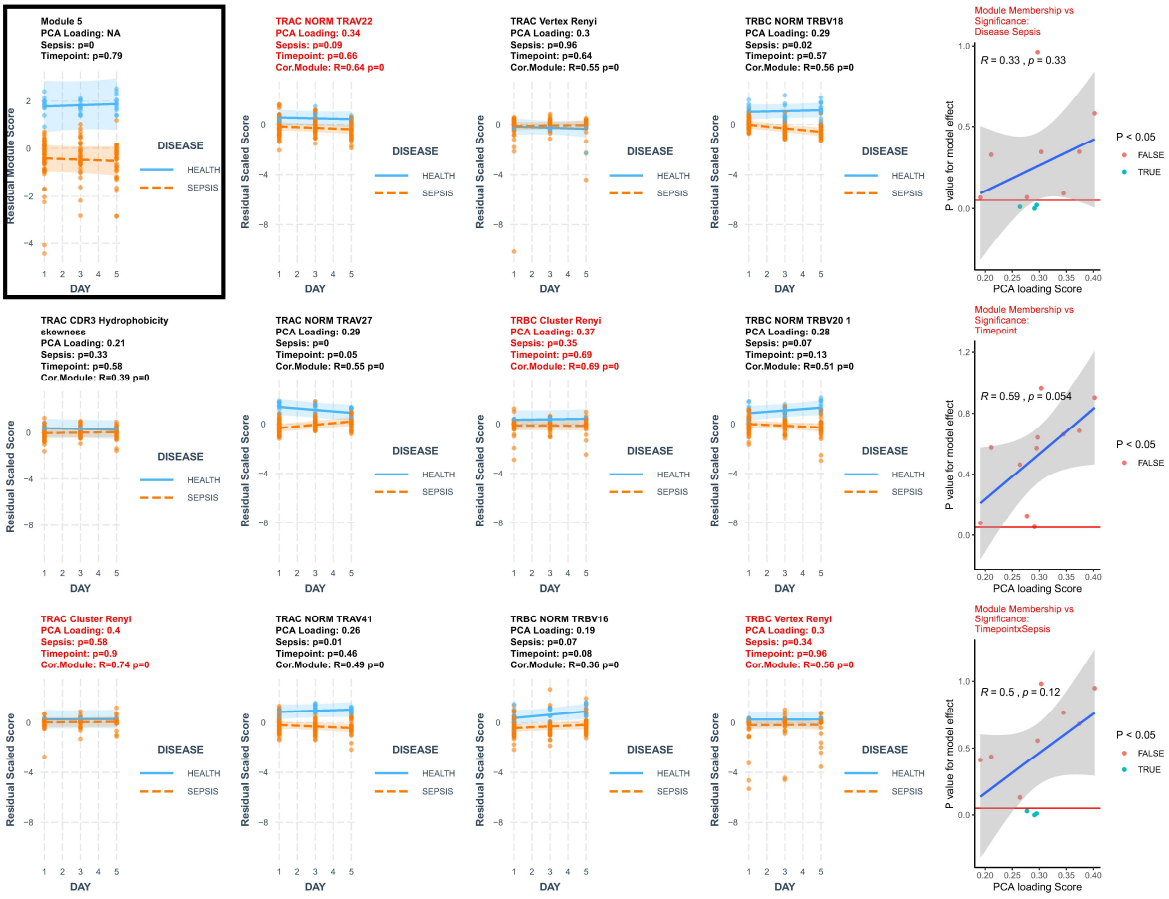


Figure 112. Partial residual plots showing the significant increase in TCRAB module 5 score in sepsis compared to health, corresponding to changes in T cell clonal expansion and diversification.

Top left shows the partial residual plot for TCRAB module 5 using the D*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 5 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). Results are reported to two decimal places.

Multiple quantitative trait associated genes were identified for modules 1, 4, and 5 [**Table 32**], but there was minimal overlap between gene sets which could be due to the relatively stringent threshold for the estimated effect size (mean ± 1 standard deviation). The genes

RTP5 and *PUS10* were strongly associated with both module 1 (RTP: ES=1.76, p_{adj}=0.034, PUS10: ES=1.88, p_{adj}=0.042) and module 4 (RTP: ES=1.72, p_{adj}=0.028, PUS10: ES=2.01, p_{adj}=0.017). The function of RTP5 is unknown, although it has been linked to highly T cell infiltrated pancreatic tumours and associated with improved outcome (Bernard et al., 2020). Similarly *AHSA1* (ES=-1.73, p_{adj}=0.0025), which was associated with module 5 , has been linked to $CD4^+$ and $CD8^+$ T cell infiltration in liver hepatocellular carcinoma (Li and Liu, 2022). *NEURL3* (Neuralised E3 Ubiquitin Protein Ligase 3) was strongly associated with module 1 score (ES=5.25, p_{adj}=$1.76e^{-08}$) and has been shown to enhance anti-tumour function in tumour-specific T cells, including increased cytokine production and upregulation of tissue resident markers such as CD69 and CD103 (Scharping et al., 2021).

Pathway analysis using XGR identified several module 1 associated pathways, including cellular response to stress, external stimuli, and CLEC7A (a pattern-recognition receptor) signalling and the innate immune system which are all plausible drivers of T cell clonal expansion [**Figure 113**]. Pathways involving the cell cycle and mitotic division were also associated with module 1, in keeping with clonal expansion. Only one pathway was associated with module 4, 'Signal Transduction' however this was not significant after BH correction. There were no pathways associated with module 5, due to the small gene-set.

Module	No. Associated Genes	Significant Genes with Largest Effect Size	Gene Overlap across Modules
Module 1	135	*AP005136.3,* ***CCDC173****, AC025449.2, AP005209.2,AC106782.3 NEURL3, AC010173.1 GTF3C1, DHFRP1, DHX38, H3C2, HSP90AA2P,* ***AHSA1****,*	**Module 4:** 2 genes (*RTP5* and *PUS10*) **Module 5:** 4 genes (*CCDC173, AHSA1, NEO1, C1orf21*)
Module 4	87	**AC013457.1**, UGT2A3 AC021483.2, SGCD, **RPA2**, CCNH, COBL INTS3, VCX, MYLK3 DGKI, **PUS10**, PTBP2 SH2D4A, **RTP5**	**Module 5:** 2 genes (*RPA2* and *AC013457.1*)
Module 5	6	***C1orf21, NEO1, RPA2, AHSA1, AC013457.1, CCDC173***	

Table 32. Genes associated with TCRAB modules 1, 4 and 5 which capture clonal expansion.

Linear mixed effects models were used to identify genes significantly associated with TCRAB module score, using bulk RNA sequencing gene expression data. P values were corrected for multiple testing using the Benjamini-Hochberg method. The number of genes significantly associated with module score (BH adjusted $p<0.05$ and effect size greater or less than the mean ± 1 standard deviation) are noted. Genes with the largest estimated effect size (descending order) are included.

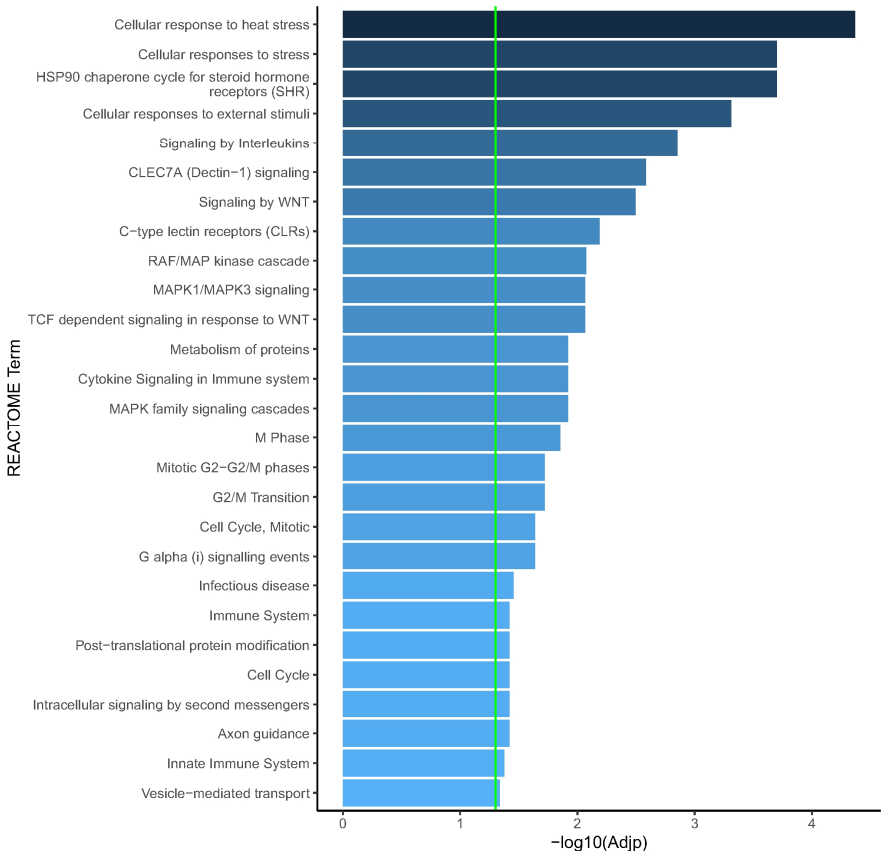


Figure 113. Pathways associated with TCRAB module 1 score, identified using XGR.

Non-AIR genes significantly associated with TCRAB module 1 score were input into the XGR tool to identify enriched REACTOME terms. The threshold of $p_{adj}=0.05$ is shown in green.

5.3.5.3 Alterations in TCRB CDR3 Length Distribution and Charge are Associated with Early-Death in Sepsis

Module 9 was the only mortality associated TCRAB module which was comprised of features other than V gene usage. In sepsis 7-day non-survivors, there was a significant decrease in module 9 score, which corresponded to a decrease in CDR3α length distribution measures (skew and kurtosis) and CDR3β charge compared to survivors. However, the mean CDR3 length did not differ **[Figure 114]**. Module 9 was negatively correlated with total white cell count (WCC), urea, and the total number of leukocytes, including both mononuclear and polymorphonuclear cells **[Figure 101]**. Despite the association with mortality it did not correlate with measures of disease severity (SAPS, APACHE and SOFA).

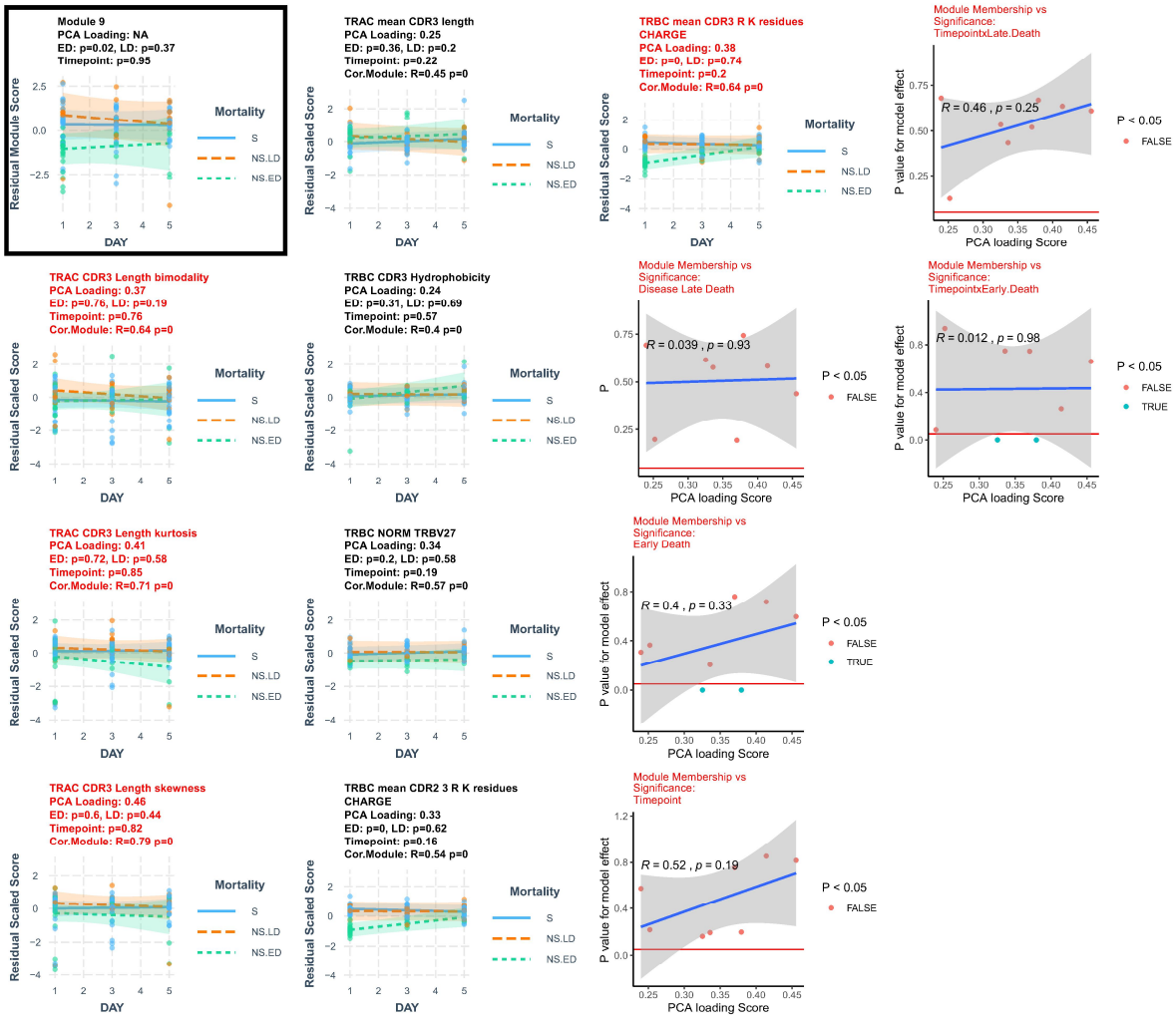


Figure 114. Partial residual plots showing the significant decrease in TCRAB module 9 score in sepsis early-death non-survivors compared to survivors and late-death non-survivors.

Top left shows the partial residual plot for TCRAB module 5 using the M*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 9 cluster, modelled using the same M*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). Results are reported to two decimal places.

In healthy individuals, the length distribution of the CDR3 regions in both TCR α and β chains typically follows a Gaussian distribution. However, during infection, this distribution can be altered (Wang et al., 2017, Ou et al., 2018, Ma et al., 2016, Kou et al., 2000). Research by H. Moss and Bell (1996) suggests that CDR3α length is less conserved compared

to CDR3β, even when transcripts with the same TCRAV gene are aligned, perhaps due to differences in P and N insertions rate. This may explain why we only detected perturbations to the CDR3α length distribution associated with mortality. Recent work by Yu et al. (2019) showed that the distribution of alpha chain lengths is more sharply peaked and less skewed than the beta chain.

The decrease in CDR3α length skew and kurtosis in early-death non-survivors indicates a 'flatter' distribution (wider-tails) of CDR3α lengths which is skewed away from larger values in comparison to survivors, similar to the changes reported in Kou et al. (2000) for individuals infected with HIV. This may suggest a less focused repertoire or increased CDR3 diversity. Several factors could contribute to the observed changes in the length distribution, including differential V gene usage, the ratio of $CD4^+$ to $CD8^+$ T cells, and alterations in palindromic and non-template insertions, which will be discussed below.

In 7-day non-survivors, several TCRAB modules (2, 7, 8, 14, and 16) were associated with mortality, but only module 17 showed a significant association with module 9 score, which includes features related to CDR3β hydrophobicity and length [**Figure 99**]. No significant correlation was found with any of the V gene mortality associated modules, indicating that changes in CDR3 length distribution may be independent of V gene usage in this cohort [**Figure 104**].

Research has shown that CDR3 regions vary in their richness, distribution and amino acid composition between $CD4^+$ and $CD8^+$ T cell populations (Li et al., 2016, Wang et al., 2017). If preferential expansion of either population was driving the change in CDR3 distribution in this cohort, you would expect an association between module 9 score and the respective

CIBERSORTx estimated cell fractions. However, this association was not observed [**Figure 104**, **Figure D. 12**].

Another possible cause for differences in CDR3 length is the insertion of palindromic (P nucleotides) and non-template (N regions) during VDJ recombination (Roth, 2014). Yu et al. (2019) have shown that nucleotide insertion and deletion positively correlate with TCR CDR3 length, shaping the CDR3 length distribution. The range of indel lengths in CDR3α regions is reportedly wider than for CDR3β, possibly reflecting a distinct role for each chain in the recognition of antigen-MHC complexes. This would be consistent with reduced conservation in length for transcripts with the same TRAV gene reported by H. Moss and Bell (1996) and explain why we only detected an association between CDR3α length distribution and mortality.

Module 9 score was associated with over 300 genes, indicating their involvement in process which determine CDR3α length or downstream responses to changes in CDR3 length [**Figure 115.A**]. Among these, *LEXM* (lymphocyte expansion molecule) had a large effect size (ES=2.28, p_{adj}=0.016). LEXM has been shown to promote antigen-dependent $CD8^{(+)}$ T cell proliferation, effector function and memory cell generation in mice by controlling the levels of oxidative phosphorylation complexes and respiration (Okoye et al., 2015). *CNOT2* had a strong negative effect on module 9 score (ES=-2.60, p_{adj}=0.03) and down-regulation of CNOT2 results in elevated expression of MHC class II genes (Rodríguez-Gil et al., 2017). Pathway analysis using XGR revealed the significant enrichment of the REACTOME term 'Immunoregulatory interactions between a lymphoid and a non-lymphoid cell', which includes genes related to MHC antigen presentation and TCR recognition. This finding is particularly interesting given the association between module 9 score and leukocyte count. Together, these results suggest that individuals with a perturbed distribution of CDR3

lengths (less focused repertoire) may have dysregulated antigen recognition. Consistent with these findings, module 9 was negatively correlated with SRSq score, indicative of immunocompromise.

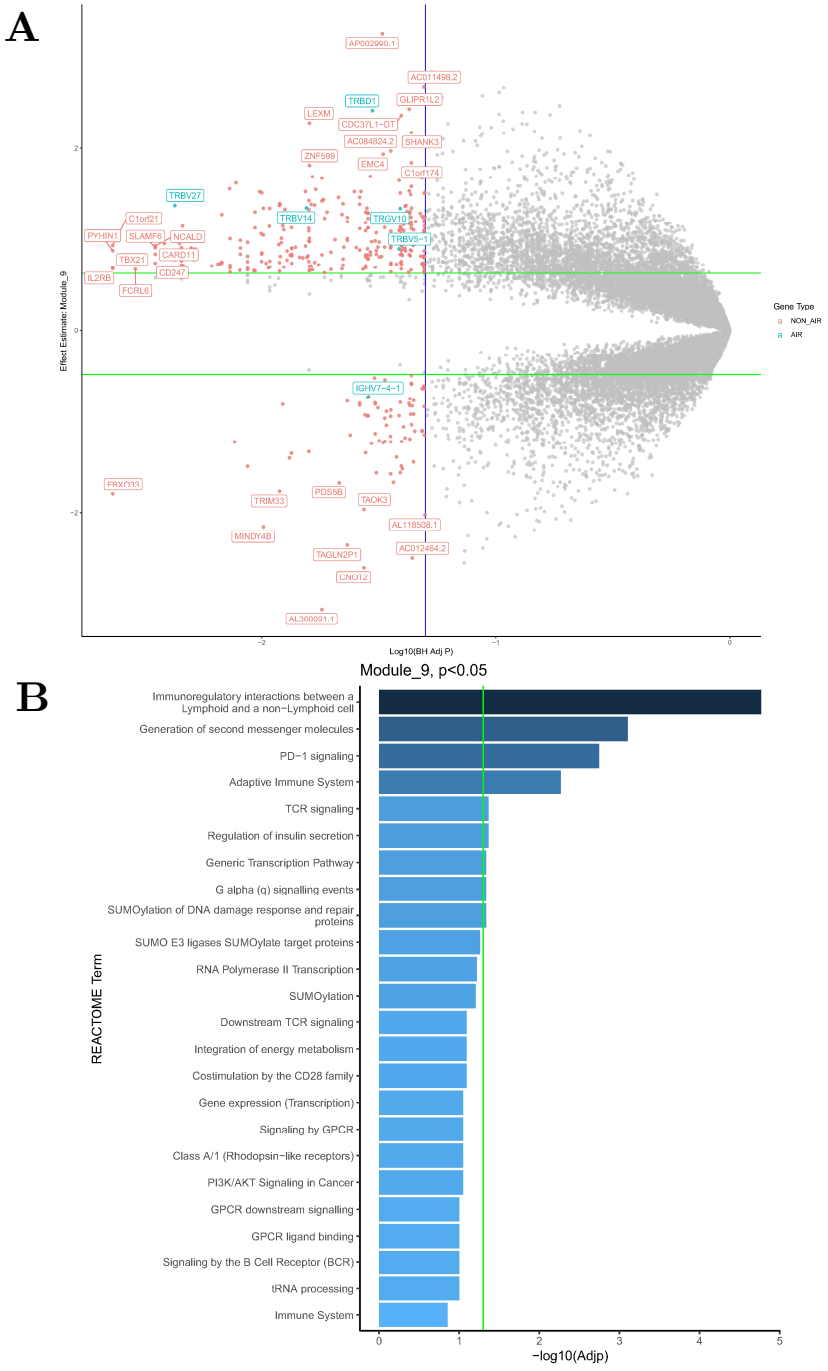


Figure 115. Module 9 quantitative trait analysis.

A) There were ~380 genes associated with module 9 score, using an ES threshold of the mean ± 1 standard deviation and $p_{adj}<0.05$. **B)** Module 9 associated non-AIR genes were input into pathway analysis using XGR. 9 REACTOME pathways were significantly enriched in relation to module 9.

5.3.5.4 Biased TRDV and TRGV Gene Usage Suggests Diversification of γδ T Cell Populations

TCRGD module 5 score was significantly increased in sepsis compared to health, capturing the normalised V gene usage for TRGV3, TRGV4 and TRGV5, as well as the proportion of unproductive TRG reads [**Figure 116**]. Module 5 showed a positive correlation with total white cell count (including mononuclear and polymorphonuclear leukocytes), but no association with organ severity scores such as APACHE II or SOFA. Similarly, module 6 was increased in sepsis and represents TRDV1 usage and mean CDR2δ and CDR3δ charge, suggesting changes in charge are driven by skewed TRDV1 usage.

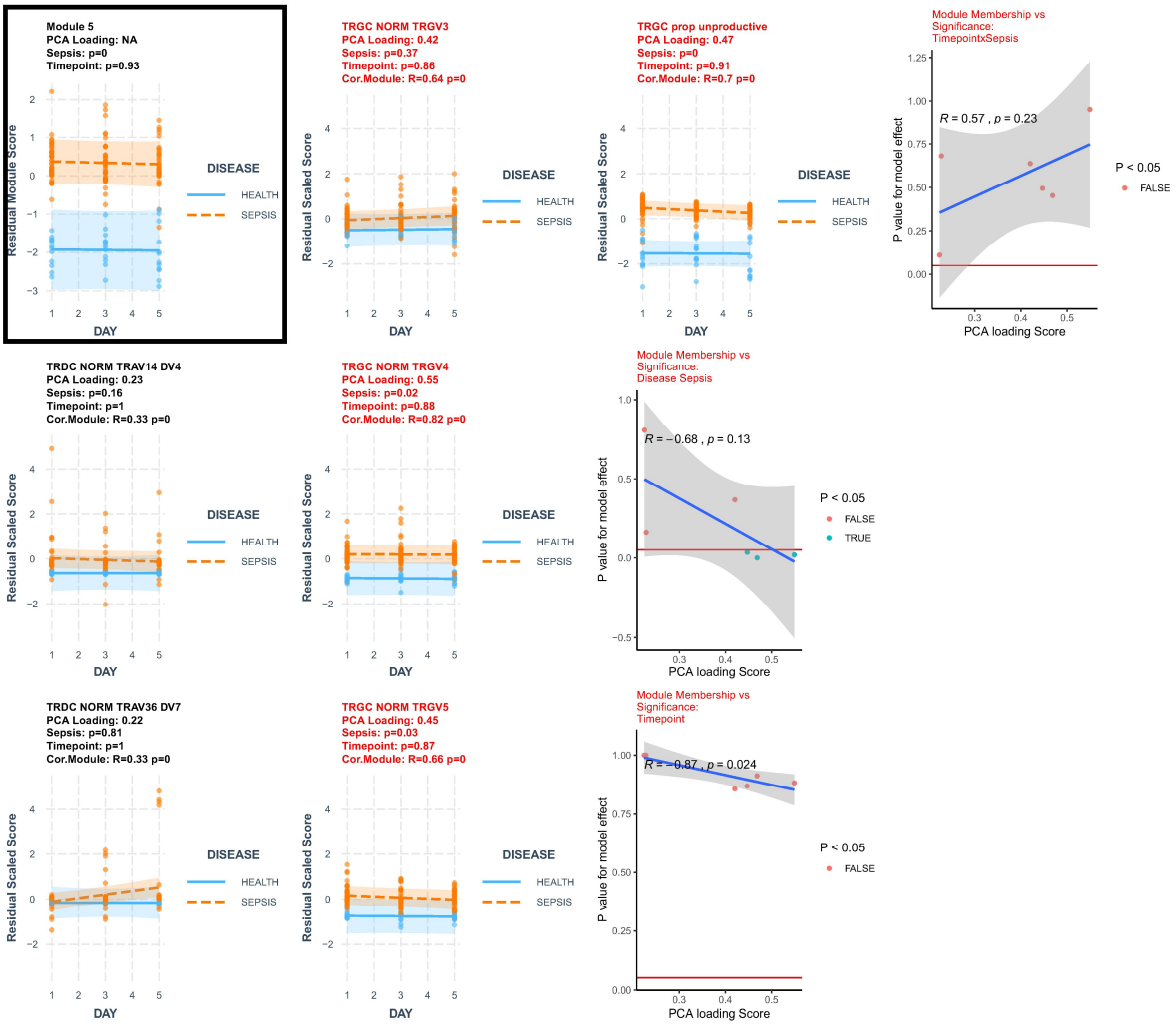


Figure 116. Partial residual plots showing the significant increase in TCRAB module 5 score in sepsis compared to health, corresponding to changes in γδ T cell subsets.

Top left shows the partial residual plot for TCRGD module 5 using the D*TASC model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 5 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). Results are reported to two decimal places.

The proportion of unproductive TRG reads is predominantly driven by the reassignment of TRGV10 VDJ sequences as 'unproductive' during data pre-processing because their origin is likely γ chain co-expression in αβ T cells. Therefore, the 'proportion TRG unproductive' feature may provide a measure of the ratio of true γδ T cells to γ-αβ T cells in bulk RNA

sequencing data. Increased score in sepsis is indicative of a reduced ratio of γδ T cells to αβ T cells compared to health. These findings align with previous studies that have reported a significant decrease in γδ T cell populations in sepsis, correlating with disease severity and that γδ T cell depletion is markedly more pronounced than for αβ T cells (Andreu-Ballester et al., 2013, Venet et al., 2005). Our results also show that the proportion of unproductive TCRG reads was significantly lower in 7-day non-survivors compared to survivors ($p<0.05$), but this could be explained by a greater decrease in αβ T cells in 7 day non-survivors [**Figure D. 16**].

Research into rare non-$V\gamma9^{+}V\delta2^{+}$ γδ T cells in humans limited. However, γδ T cells which express $V\gamma4^{+}$ have been shown to localise to the intestinal epithelium in humans and possess an innate-like (semi-invariant) phenotype in humans. They are regulated by both BTNL3 and BTNL8 via the CDR2 region; an interaction akin to super antigen-like binding and also by clonally restricted antigen via CDR3 mediated recognition (Fichtner et al., 2020, Willcox et al., 2019). Murine Vγ5 γδ T cells exclusively locate to the skin epidermis where they exist in contact with keratinocytes and may function to repair damaged epithelial tissues (Boismenu and Havran, 1994, Shepherd and McLaren, 2020).

The subset of γδ T cells which express Vδ1 have been shown to have potent anti-tumour functions and respond to conserved signals of tissue stress via innate receptors. They are the dominant subtype in lung tissue (compared to the blood where most γδ T cells express Vδ2) and are also prominent in adult skin, intestine, liver, mucosal tissues and the colon epithelia where they recognise antigen from virally infected cells and play an important role in immune-surveillance (Wu et al., 2022a, Pizzolato et al., 2019, Bruni et al., 2022). Recent work has shown that unlike $V\gamma9^{+}V\delta2^{+}$ T cells, $V\delta1^{+}$ populations are able to bind underneath the antigen-binding cleft of the monomorphic MHC-related protein 1 (MR1)

which is known to present vitamin B precursors and by-products to mucosal associated invariant T (MAIT) cells. These Vδ1 MR1-restricted γδ T cells populate damaged tissues, indicative of a role in pathology (Le Nours et al., 2019).

The alterations in TRGV gene usage observed in sepsis are therefore likely to reflect the need for diversification of the γδ T cell response, to target localised infections in lung tissue, disseminated infections and tissue damage. Interestingly, elevated Vδ1 usage in non-survivors (early and late-death) compared to survivors suggests a potential role as a biomarker of severe disease and is consistent with previous reports in the sepsis literature (**5.1.1.2**).

5.3.5.5 γδ T Cell Clonal Expansion and Loss of Diversity

TCRGD module 1 was significantly elevated in sepsis 7-day non-survivors compared to survivors and included TCRD clonal expansion measures (maximum vertex size and the proportion of the repertoire explained by the top 5, 10 and 50 clones) which provide a measure of repertoire diversity. A similar association with mortality was seen for TRG (module 2) though this did not reach statistical significance – perhaps because the delta chain is more diverse than the gamma, due to the diversity (D) gene segment, which allows for more accurate clonal detection (Hovav, 2017). Reassuringly, TCRG and TRD clonal expansion measures are closely correlated, despite separate PCR amplification [**Figure 100**].

Both module 1 and 2 were elevated in infection with *Streptococcus pneumoniae,* therefore it would be interesting to compare module score between CAP and FP sepsis to determine whether clonal expansion is unique to respiratory infection or a more general marker of mortality. There was no association with non-AIR genes for either module 1 or module 2.

5.3.6 Integrating TCR and BCR Repertoires

To understand high-level organisation of the BCR and TCR repertoire in sepsis, RMCORR was used to correlate BCR, TCRAB and TCRGD modules [**Figure 117**]. There were very few significant associations ($p<0.05$) and none remained significant after BH multiple testing correction, perhaps due to the large number of non-independent tests performed.

There was a strong positive correlation between TCRGD module 1 and TCRAB module 1 which capture TCRD clonal expansion and TCRA clonal expansion respectively, suggesting that both populations behave similarly within an individual in the context of sepsis [**Figure 117.C**]. However interestingly, BCR clonality (module 19 and inverse module 13), including IgG1 clonality (BCR module 31) was negatively correlated with TCRGD clonal expansion (TCRGD module 3, 10 and 11). This relationship may reflect a decrease in the conventional semi-invariant $V\gamma9^{+}V\delta2^{+}$ T cell subset (module 4) which are the majority of γδ T cells health and typically display an oligoclonal repertoire, in favour of diversification of the γδ T cells response with the more adaptive-like Vδ1− γδ T cells during sepsis, which recombine with a more diverse set of V genes (**5.3.5.4**) (Clark and Thomas, 2020). Or alternatively, mobilisation of γδ T cells into the tissue.

Gamma-delta T cells can be found inside germinal centres, where they influence B cells through cell-to-cell contacts and indirectly through regulation of T_{fh} cells. Research in mice has shown that depletion of certain γδ T cell subsets alters total Ig levels in sera, however this was not observed, perhaps due to differences between immunoglobulin gene expression and serum protein levels, because of the relative contribution of different B cell types (Rampoldi et al., 2020).

When comparing between the BCR and TCRAB repertoire, the majority of associations were between BCR and TCRAB V gene containing modules, which suggests the expansion of B and T cell clones which recognise antigen and processed presented antigen (p-MHC complexes), from the same origin, for example TCRAB module 2 and BCR module 25 (IGHV2 usage). Interestingly BCR IgG SHM (module 9) had a strong positive correlation with TCRAB module 10, which captured changes in the distribution of TCRB CDR3 length, which may again reflect response to the encountered antigen.

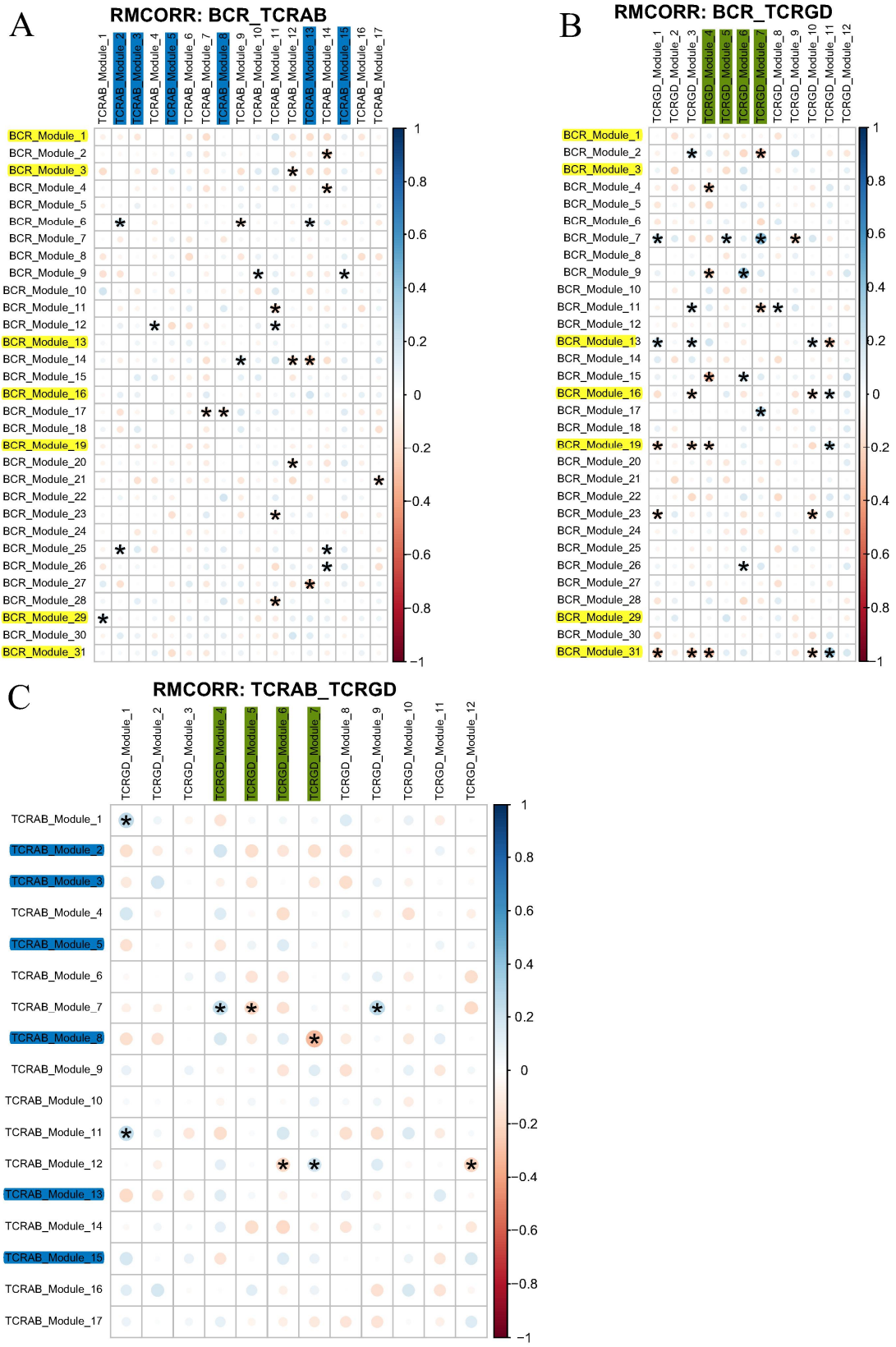


Figure 117. Correlation between TCR and BCR modules, estimated using RMCORR.

Significant correlations ($p<0.05$) are indicated with an asterisk. Modules which were significantly different between health and sepsis are highlighted. **A)** BCR versus TCRAB modules, **B)** BCR versus TCRGD modules and **C)** TCRAB versus TCRGD modules.

5.4 Discussion

5.4.1 TCR Repertoire Sequencing: Technical Challenges

In this chapter, I have conducted an in-depth analysis of the TCR repertoire in sepsis. To ensure adequate read depth, each TCR chain was amplified separately using a multiplex PCR reaction. This approach was important for capturing rare γδ T cells, which are significantly depleted in sepsis. TCR repertoire sequencing depth is influenced by biological factors such as target T cell count, as well as technical factors including RNA input quantity, PCR efficiency, number of multiplexed samples and off-target products. Consequently, a limitation of this study is the inability to directly compare the read depth of α, β, δ and γ chains and estimate the ratio of αβ to γδ T cells, which could potentially serve as a useful marker for mortality in sepsis. However, the ratio of TRGV10:TRGV9 usage might serve as a proxy due to γ-αβ co-expression, although further validation is needed.

The module scores for the majority of TCRAB and TCRGD modules were stable in healthy individuals during the 5 day period. However, in some cases, there were statistically significant longitudinal changes in module score, most likely due to the substantial differences in sequencing depth from errors in one-to-one pooling of libraries [**Figure 81**]. This was predominantly observed in modules related to V gene usage and clonal expansion. Therefore, I refrained from making conclusions regarding longitudinal dynamics in these modules and findings should be interpreted with caution. Although the differences between health and disease in these modules was greater than the longitudinal variation within each group, suggesting genuine disease perturbations.

V gene usage was calculated as a percentage of the total repertoire; metrics were not calculated using repeated subsampling, under the assumption that they should be independent of read depth. However, in practicality, when there was a large difference in read depth between longitudinal samples, or low diversity, a minimal sample depth is required to detect rarer V genes and obtain a stable estimate. For the BCR repertoire, samples with a productive read depth less than 1500 were excluded, the resulting BCR modules were stable in health. However, for TCRAB and TCRGD, samples were excluded based on a productive read depth of <300 and <50 respectively. This threshold was likely too low to capture rarer V genes and may explain the variability seen in V gene modules. However, this threshold was a compromise to avoid sample dropout, as I aimed to exclude no more than 5-10% of samples. A more comprehensive approach to determine this read depth threshold could be to perform a rarefaction analysis, using number of unique V genes detected as a function of subsample depth (instead of the number of unique VDJ sequences). The average read depth at which V gene detection plateaus across samples could be used as a threshold for filtering, resulting in more stable estimates of V gene usage.

Regarding clonal expansion measures, which are down-sampled to reduce sample depth-associated variation, the instability in TCRGD module 2 score in healthy controls across time, also highlights that large differences in read depth, where one sample has relatively few reads (<1500) cannot be fully corrected for, even following best practices. Therefore, utmost care should be taken when pooling samples for sequencing. In the TCR experimental protocol there was significant off-target binding and primer-dimer formation, likely due to the large number of PCR primers involved, hence this process was challenging. As noted earlier, the use of double-sided strand selection during library preparation should yield a purer product and increase sequencing depth, helping to address this issue.

Like the BCR repertoire, a disadvantage of unpaired TCR sequencing is the inability to characterise the functional properties of T cells, as antigen specificity is mediated by both chains. This may be of less importance for the γδ T cell repertoire where the majority of T cells are $V\gamma9^{+}V\delta2^{+}$ and exhibit a semi-invariant repertoire compared to the αβ T cell repertoire.

5.4.2 Sepsis is Characterised by Subtle Alterations to the TCR repertoire

5.4.2.1 Sample Size

A recurring observation throughout the TCR repertoire analysis, was small effect sizes across multiple similar repertoire features when comparing between health and sepsis or mortality groups. These changes were not always statistically significant, in contrast to the BCR repertoire where more pronounced perturbations were observed. However, this clearly highlights the benefits of using a dimensional reduction pipeline such as rWGCNA, which captures the key axes of variation across multiple features and consequently is able to separate health and disease. Similarly, projection of PC1 and PC2 coordinates clearly discriminates between health and sepsis for both the αβ and γδ TCR repertoire demonstrating that overall there is profound alteration in the sepsis repertoire (**5.3.4.7**). The lack of statistical significance could be explained by a variety of factors.

Firstly, although our cohort initially consisted of 72 sepsis patients and nearly 200 longitudinal samples, the generally lower read depth meant fewer samples passed filtering criteria compared to the BCR repertoire. Furthermore, by combining data from paired chains into a single matrix, a small number of samples with data available for only one chain were excluded. Collectively, this reduced cohort size, and may mean the study was

underpowered to detect such small changes in the repertoires. Analysis could be repeated for each TCR chain separately to improve statistical power. However, even with this approach, validating the results through cross-receptor correlations would still be necessary as measures of clonal expansion are expected to be correlated within each receptor, and would require the subset of samples with matched data.

5.4.2.2 Confounding Variables

Results may also be confounded by cohort heterogeneity. There were multiple associations between module score and factors such as infectious agents (*e.g. Streptococcus pneumoniae*) and treatment with Drotrecogin alfa or corticosteroids. These covariates were not included in the original model to assess the effect of disease status or mortality group and time on module score [**Figure 101**, **Figure 102**]. Failing to include these significant sources of variation could decrease the significance of main effect terms.

Interestingly there was a positive correlation between both TCRB and TCRG clonal expansion modules and treatment with corticosteroids. This finding aligns with studies conducted in murine models, which suggest that dexamethasone-induced apoptosis primarily affects naive and memory $CD8^+$ T cells, while virus-expanded effector $CD8^+$ T cells are more resistant to apoptosis (Kumar and Sehrawat, 2019). Additionally, Azeez et al. (2022) show that transient dexamethasone therapy can lead to the expansion of virus-specific $CD8^+$ T cells, which preferentially migrate to infected tissue sites. Therefore, the correlation between corticosteroids and clonal expansion could be explained by a larger proportion of the repertoire being comprised of preferentially retained virus-expanded $CD8^+$ T cells in treated individuals. However, considering that these modules also correlated with markers

of disease severity, it is possible that this association reflects disease pathophysiology rather than treatment.

There was a strong signal between alterations to the TCRAB repertoire (7 modules) and potassium levels measured at day 1, post-ICU admission, including a positive correlation between potassium levels and αβ T cell clonal expansion (modules 1, 4 and 5). This is particularly interesting given no such associations were seen for any TCRGD or BCR modules. In the tumour microenvironment, infiltrating lymphocytes are dysfunctional and often exist in a quiescent state next to growing tumour cells, however in some patients T cells retain the ability to proliferate, persist and mediate destruction of the tumour deposits. Potassium is found at high levels in the tumour microenvironment due to its release by necrotic cells and this leads to the reduced uptake and consumption of nutrients by resident T cells promoting a starvation response which is characterised by autophagy and reduction of histone acetylation at effector and exhaustion loci. T cells are preserved in a stem-like state, retaining the ability to divide but have supressed effector function. This in turn produces $CD8^+$ T cells with improved in vivo persistence, multi-potency, and tumour clearance (Vodnala et al., 2019). Interestingly, the authors note these cells are transcriptionally similar to $CD8^+$ T cells which retain stemness and the capacity to respond to blockade of the PD-1 inhibitory therapy in the context of chronic antigen exposure (Vodnala et al., 2019, Eil et al., 2016, Im et al., 2016). Unfortunately serum potassium was not measured at each time point, which would have increased the power of this analysis.

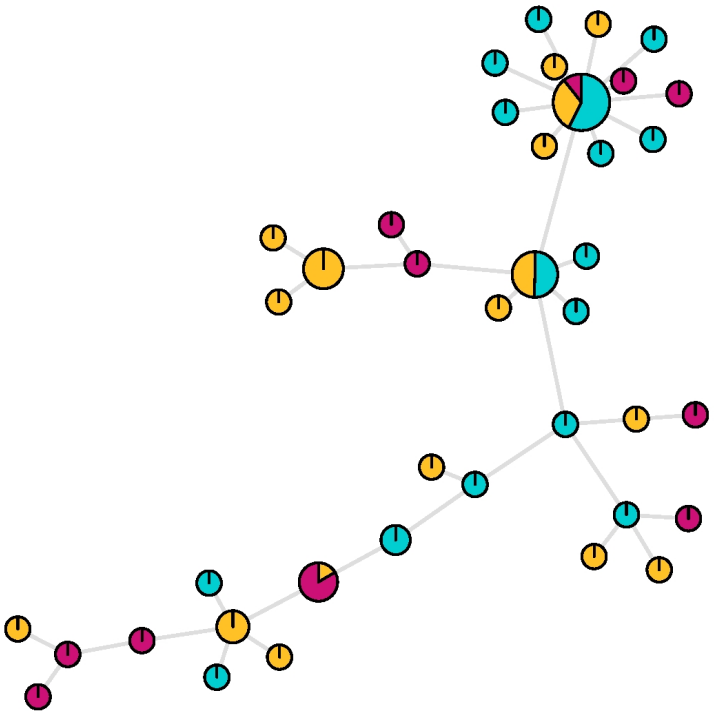
5.5 Conclusion

In this chapter I have demonstrated the wider applicability of the rWGCNA pipeline by characterising the αβ and γδ T cell response to sepsis, during the five days post-ICU

admission. Although the majority of perturbations to the repertoire were subtle, I have identified several alterations from health including increased clonal expansion and changes in V gene usage which correlate with clinical variables and disease outcome. Perturbations to TCRG/D V gene usage are particularly informative as they are associated with T cell localisation and function, and provide evidence of diversification of γδ T cell response. By comparing module score to clinical variables, I provide evidence of pathogen-specific alterations to the TCR repertoire, which suggest targeted immune response and is an interesting foundation for further analysis.

In a novel computational approach, I integrated TCR repertoire modules with gene expression data and estimated cell fractions and identified potential mechanisms involved in shaping the TCR repertoire and downstream effector functions. Furthermore, by profiling both the TCR and BCR repertoire, commonality and differences in the B and T cell response to sepsis were described.

Chapter 6. Discussion and Future Directions



6 Discussion and Future Directions

This chapter outlines the broader conclusions and limitations of the work presented in this thesis and describes areas for future research.

6 Discussion and Future Work 363

6.1 Overall Summary: The Adaptive Immune Repertoire in Sepsis 363

6.2 Discussion and Limitations 365

6.3 Future Directions 375

6.4 Conclusion 381

This thesis has, for the first time, characterised the dynamics of both the BCR and TCR repertoire in a cohort of CAP sepsis patients, over the five days following admission to ICU. I have employed an optimised in-house high-throughput multiplex PCR reaction, coupled with next generation sequencing to amplify the BCR IgH, and TCR α, β, γ and δ chains independently. As part of this thesis, I developed a novel dimensional reduction analysis pipeline for repertoire data (rWGCNA), which I have applied to the sepsis data to identify features of the adaptive immune repertoire associated with the sepsis response, specifically mortality, SRS score and clinical covariates. In addition, I have integrated rWGCNA modules with bulk RNA sequencing data to identify genes associated with perturbations to the adaptive immune repertoire. This chapter summarises the findings of this thesis, outlines the general limitations of the study and suggests areas for future research.

6.1 Overall Summary: The Adaptive Immune Repertoire in Sepsis

The GAinS study provides researchers with a valuable resource of well-characterised longitudinal sepsis patient samples. I used a subset of these patients (n=72) for comprehensive sequencing of the adaptive immune repertoire, in the week post-ICU admission, focusing specifically on Caucasian individuals with sepsis due to CAP, to reduce

cohort heterogeneity and increase statistical power. This enabled the detection of small changes in the adaptive immune repertoire at a level not achieved by previous studies. To our knowledge, this study is the first to perform BCR IgH, αβ and γδ TCR repertoire sequencing on a matched cohort of sepsis patients, providing insights into the relationships between different biological compartments.

In terms of methodology, this thesis offers comprehensive start-to-finish guidelines for analysing the adaptive immune repertoire. I address wet-lab PCR amplification, computational pre-processing, quality control, generation of repertoire summary measures and demonstrate the use of technical replicates for batch effect correction. Of particular importance in a field where detection of public sequences is a frequent goal for analysis, I describe a method of quantifying library index hopping and contamination. I also discuss the unique challenges associated with TCRG repertoire amplification, due to γ-αβ expression.

In **Chapter 3**, I developed and optimised a novel computational pipeline based on WGCNA for dimensional reduction and identification of co-regulated features in BCR and TCR repertoires. This has the advantage of reducing the multiple testing burden, and increasing power to detect significant differences between groups of patients. This work will benefit the wider AIR community, and complements the commonly used PCA approach. Importantly, I address the implications of unbalanced longitudinal study design on correlation-network analysis; an extension to the work presented by Li et al. (2018).

In **Chapter 4**, I characterised a variety of changes to the BCR repertoire in sepsis, including alterations in immunoglobulin isotype usage, increased IGHA1 to IGHA2 CSR, clonal expansion and the presence of a putative IgD^{+} plasma population associated with disease

severity. Many of these features can be linked through a review of the literature to respiratory infection, consistent with our choice of CAP sepsis patients. By applying aspects of network analysis I identified biologically relevant changes in B cell clone structure between health and sepsis.

In **Chapter 5,** I characterised the changes to the αβ and γδ T cell repertoire in sepsis. Similar alterations were observed, including clonal expansion, biased V gene usage and the diversification of the γδ T cell response. However, these were typically less pronounced than for the BCR repertoire and did not always achieve statistical significance. In these cases, dimensional reduction approaches were vital to fully delineate health and disease.

By integrating repertoire modules with bulk RNA sequencing data, I validated repertoire signatures using a separate methodology, and identified positive correlations between rWGCNA modules including features such as clonal expansion and the estimated fraction size of the relevant cell populations. This analysis also identified module-associated pathways enriched with genes relevant to the adaptive immune response including mechanistic pathways *e.g.* those involved in the increased proliferation and metabolism required for rapid clonal expansion and effector functions such as immune-regulatory interactions between lymphoid and non-lymphoid cells. This novel analysis may be a foundation for future biomarker discovery in immune-mediated and infectious diseases.

6.2 Discussion and Limitations

There are various limitations to the work described in this thesis, some of which were discussed at the end of the relevant chapter. The following sections will summarise the major points applicable to my study of the adaptive immune repertoire in sepsis.

6.2.1 Adaptive Immune Repertoire Stability in Health

To our knowledge, this is the first study to perform BCR and TCR repertoire sequencing on sequential healthy volunteer samples within a short time-frame (5 days). Typically, studies interested in longitudinal adaptive immune repertoire dynamics, such as in the context of vaccine immunisation or treatment for auto-immune disease, identify time-dependent alterations by comparing across patient subgroups *e.g.* treatment versus placebo (Laserson et al., 2014, de Bourcy et al., 2017, Bashford-Rogers et al., 2013). The focal longitudinal groups may be compared to a healthy donor cohort, but often only at a single 'baseline' time point (Niu et al., 2020a, Galson et al., 2015b). In some studies, no healthy controls are included (Niu et al., 2020b, Khatri et al., 2021).

Several projects have specifically investigated the longitudinal adaptive immune repertoire in health. Mitsunaga and Snyder (2020) investigated the BCR repertoire in a single patient at 24 time points over the span of 11 months, which included two instances of natural infections. Results indicated that IGHV gene usage is stable across time, regardless of health status. Similarly, Sun et al. (2022) characterised the TCR repertoire in 30 healthy participants in two samples separated by an average of 9 years. The authors report only a minimal reduction in TCRα and TCRβ repertoire richness for both naive $CD4^+$ and $CD8^+$ T cells during this time. Thus, the limited literature suggested that minimal biological variation in the healthy volunteers' repertoires over five days should be expected. However, variation may be introduced by frequent blood sampling, which in patients in critical care (averaging around 30mls a day) contributes to iatrogenic anaemia after just three days (Lennox and Bench, 2022). Technical factors such as re-amplification and sequencing and finite read depth will also contribute to variation in the adaptive immune repertoire. Including healthy temporal control samples in our study was therefore necessary to

distinguish between dynamic immune responses and technical variations in module score during acute disease.

In agreement with the literature, every BCR repertoire module was stable in health, hence I am confident that the subset of modules that showed a significant change in module score through time in sepsis reflect dynamic immune processes. The healthy control dataset will be valuable for researchers interested in profiling acute disease dynamics and understanding the longitudinal adaptive immune repertoire. It provides additional evidence of the temporal stability of the adaptive immune repertoire, in agreement with previous publications.

In the case of the TCR repertoire, although the majority of modules were stable in health, there were a small number with a statistically significant change in module score across time. As discussed, in **Chapter 5,** this is likely due to substantial differences in read depth between samples and a low minimum read depth required for subsampling.

6.2.2 Time of Sepsis Onset

The detection of several dynamic BCR modules in sepsis is particularly interesting. Upon infection, adaptive immune system activation is delayed by 4-10 days and increased clonal overlap and B cell clonal lineage extension typically takes 2-3 weeks to establish (Zheng et al., 2022, Riddell). For instance, in SARS-CoV-2 infection, massive clonal expansions were observed in the second week after symptom onset, although a transient surge in IgA was found in the first week (Zhang et al., 2022). Given that many BCR and TCR modules were elevated in sepsis compared to health at time point 1, including BCR module 3 (IgD/M SHM) and BCR module 13 (clonal expansion and diversification), it suggests either the presence of pre-existing differences that predispose to sepsis, or, more likely, that patients have been infected for a significant period of time prior to ICU admission. This second

explanation is consistent with work by both Liu et al. (2018) and Jacob et al. (2023) which found that over 33%-45% of sepsis patients had clinician-based encounters in the week prior to hospitalisation, with an increased incidence of diagnosis of acute infection and antibiotic use. Neilson et al. (2023) report that in a substantial number of cases, patients make as many as 2-4 clinician visits within 72 hours of hospitalisation with sepsis.

Therefore one of the major limitations of this study is the differing time of disease onset between patients, which confounds results. There is some evidence of this in my analysis. A subset of CSR frequencies (typically in IgG isotypes) were significantly elevated in sepsis early-death and late-death non-survivors compared to health at day 1. The significant elevation relative to health was not seen for sepsis 6-month survivors until later time points (day 3 and 5). This may indicate that non-survivors had delayed diagnosis and admission, hence infection and adaptive immune activation had progressed further prior to ICU admission. This observation aligns with the literature, which has found that early detection of sepsis and appropriate intervention increases survival rate (Kim and Park, 2019).

Similarly, modules that changed significantly with time, were typically restricted to sepsis survivors for the BCR repertoire. While this may be an artefact from the loss of statistical power due to sample drop-out in the non-survivors group, it could also reflect the earlier stages of adaptive immune activation in survivors. In future, it would be beneficial to investigate adaptive immune response to sepsis in a cohort of patients where the time of infection is known, such as post-operative surgery, so that trajectories between sepsis survivors and non-survivors (and those who do not develop sepsis) better align (Mulita et al., 2022).

In many cases, module score did not return to baseline health during the five days. Previous studies in humans with sepsis have shown normalisation of the adaptive immune repertoire after sepsis, by the later time points of day 7 to day 30. Therefore, it would also be interesting to expand the sampling timeframe and sequence the repertoires of ICU discharge and convalescent samples (Venet et al., 2013). This would allow us to investigate sepsis induced accelerated immunosenescence, and susceptibility to recurrent infection. Unfortunately, recruitment of convalescent samples for this study was prevented by restrictions during the COVID-19 pandemic, however they are now a focus of ongoing recruitment and research by the Sepsis Immunomics study (an extension of GAinS).

6.2.3 Infectious Aetiology: Community-Acquired Pneumonia versus Faecal Peritonitis

During the analysis of the BCR and TCR repertoires in **Chapters 3 and 4**, I consistently observed changes in sepsis that align with previous studies characterising changes to the adaptive immune repertoire in various respiratory infections, consistent with our choice of CAP sepsis patients. Sepsis is the dysregulated response to infection and is believed to occur via a common biological pathway. However, it is acknowledged that the pathogen involved can influence treatment choice and outcome. Therefore, given the specificity of the adaptive immune response, it is not surprising that associations between module scores and infection with specific pathogens such as *Streptococcus pneumoniae* or *Haemophilus influenza* were detected, particularly in the case of the γδ T cell repertoire as these cells are known to be involved in resolution of common pneumonia causing pathogens (Kirby et al., 2007, Cheng et al., 2012, Ivanov et al., 2014). Gaining an understanding of both the shared and pathogen-specific adaptive immune responses will be crucial for the development of more targeted therapies for sepsis. This is particularly challenging given the limited microbiology

information available for these patients, meaning many may have undiagnosed infection. Ongoing work as part of GAinS includes expansion of patient microbial data, for example through digital droplet PCR and clinical metagenomics, which should, in time, increase the power to compare host responses to different infectious aetiologies (Goh et al., 2020).

The GAinS study has also recruited patients with sepsis caused by faecal peritonitis (FP), defined as the bacterial seeding of the peritoneal cavity from faeces due to disruption of the integrity of the gastrointestinal tract (Clements et al., 2021). This commonly occurs as a result of trauma or surgical perforation of the large bowel. Mortality for sepsis from FP is as high as 40% in Britain (Weledji and Ngowe, 2013). Unlike CAP, FP sepsis is usually poly-microbial in origin, and associated with a high bacterial load. Common causative agents include *Streptococcus spp*, *Escherichia coli* and *Enterobacteriaceae* (Eckmann and Bassetti, 2014). Previous research on the GAinS cohort and additional validation groups has shown that while the transcriptomic response to CAP versus FP sepsis is largely independent of the source of infection (Burnham et al., 2017), there are a small number of genes associated the infectious origin. For instance, CAP patients show enrichment in pathways related to interferon (IFN) signalling and antigen presentation, with significant upregulation of the genes: *IFIT1*, *IFIT2*, and I*FIT3* (IFN-induced antiviral genes), *BTN3A*3 (major histocompatibility complex class I gene), *IFIH1* (sensor of viral nucleic acids), and *OAS2* (IFN-induced antiviral enzyme) which indicate a prominent viral signature (Burnham et al., 2017). Given these distinct transcriptomic differences, particularly in antigen presentation, it is reasonable to expect variations in the adaptive immune repertoire response between CAP and FP sepsis which could be a premise for future research.

6.2.4 The Dysregulated Immune Response and Choice of Controls

Distinguishing between the *dysregulated* (maladaptive) sepsis immune response and the *normal* appropriate immune response to infection, which leads to pathogen clearance and restoration of homeostasis, can be challenging without appropriate controls. This is made more difficult by the diverse infectious aetiology of sepsis and use of disease cohorts consisting of older individuals with significant age-related health conditions and comorbidities. In this study, 12 age and sex matched healthy donors without underlying health conditions or acute infection were recruited. Recruitment was challenging given the requirement for three blood samples within a 5-day timeframe, resulting in low participation rates.

Analysis of the TCR repertoire suggests that this study may lack sufficient statistical power to detect subtle changes in the adaptive immune repertoire during sepsis, potentially exacerbated by the small number of healthy controls. As the repertoire of healthy individuals was generally stable over time (**6.2.1**), including additional healthy controls sampled at a single time point could enhance the statistical power of this study.

To account for significant pre-existing patient comorbidities that impact long-term mortality in sepsis patients, the Charlson Index was used as a covariate in statistical analysis. There were insufficient numbers of patients with each comorbidity to warrant an in-depth statistical comparison. However, the use of the Charlson Index fails to consider the heterogeneity in comorbidities and the specific impact of each condition on the adaptive immune repertoire. For example, two sepsis patients with Rheumatoid Arthritis (R) were included in the sepsis cohort, but previous comprehensive profiling has revealed disease-specific perturbations to the BCR and TCR repertoire, such as T cell oligoclonality in RA

(Lo et al., 2020, Goronzy et al., 1998). Thus heterogeneity in comorbidities may confound results and reduce statistical significance.

In an ideal study, it would be beneficial to include non-sepsis controls with severe or mild respiratory infection, as well as non-infectious inflammatory controls. For instance, in a sepsis study conducted by Burnham et al. (2017), a control group of patients undergoing elective cardiac surgery, who experienced sterile inflammation, was used. Similarly the COVID-19 Multi-Omic Blood Atlas (COMBAT) used sepsis patients from GAinS, as a control group for assessing COVID-19 mild and severe infection (Ahern et al., 2022). This will be further discussed in section **6.2.4.1**. Due to time constraints imposed by the COVID-19 pandemic, control recruitment was limited to healthy volunteers. Hence, cross-disease comparisons were restricted to within-sepsis subgroups, such as SRS and mortality groups. In the future, a meta-analysis could be performed to compare sepsis with other severe respiratory infection cohorts, such as hospitalised versus non-hospitalised patients with *Haemophilus Influenza*, to understand how the adaptive immune response in sepsis is dysregulated.

6.2.4.1 The COVID-19 Multi-Omic Blood Atlas

During the COVID-19 pandemic (2020), the Bashford-Rogers and Knight Groups were part of the COVID-19 Multi-Omic Blood ATlas (COMBAT) Consortium and conducted BCR and TCRβ repertoire amplification and data pre-processing using the methods described in this thesis. Patient samples were collected from individuals with mild, critical and severe COVID-19 infection (Ahern et al., 2022). Additionally, 58 newly recruited all-cause sepsis patients were included as a comparison group, due to the similarities between sepsis and severe COVID-19 infection, such as cytokine hyper-activation and lymphocyte exhaustion.

Although the cause of sepsis in COMBAT is more variable than those presented in this thesis and only one time point was included, it provides valuable context for interpreting some of the findings in this thesis.

Key findings from COMBAT revealed that COVID-19 infection was characterised by significant plasmablast expansion, which was associated with disease severity. However, in sepsis there was no significant expansion compared to health, indicating an impaired/dysregulated response. In the larger longitudinal study described in this thesis, significant clonal expansion was observed in sepsis compared to healthy individuals. This discrepancy can be attributed to two factors, the increased study size allowing for greater statistical power, and differences between sepsis patients with community-acquired pneumonia (CAP) and those with sepsis of other causes.

The mean level of SHM in plasmablasts was significantly higher in sepsis compared to mild but not severe COVID-19 infection. This finding is consistent with our observation of highly mutated IgM and IgD sequences, which were associated with mortality. Collectively, these results suggest that SHM could serve as a marker of disease severity in infectious disease (Ahern et al., 2022). Sepsis patients had significantly fewer unmutated IGHV4-34 AVY/NHS motifs compared to patients with mild and severe COVID-19. This is particularly intriguing, given un-mutated IGHV4-34 AVY/NHS are autoreactive but bind commensal bacteria (Schickel et al., 2017). Additionally, differential expression of many immunoglobulin IGHV genes was observed between COVID-19 and sepsis.

Consistent with previous studies, the main findings of bulk TCRβ sequencing and single cell VDJ sequencing in COMBAT were reduced TCR diversity in sepsis and a lower proportion mucosal-associated invariant T (MAIT) cells. Interestingly, these measures were similar

between sepsis and severe COVID-19 and the difference from healthy controls was more pronounced than in mild COVID-19 infection.

The results from the COMBAT study highlight both similarities and differences between sepsis and severe and mild COVID-19 infection, initial evidence of a dysregulated response in severe infection. In the future, the dataset generated by COMBAT could be integrated with the TCRβ/BCR data in this thesis, allowing for a larger meta-analysis given the identical experimental methods and computational pre-processing.

6.2.5 Racial Disparity in Sepsis

The majority of patients recruited from UK ICUs as part of GAinS have been Caucasian [**Figure 11**]. In an effort to reduce cohort heterogeneity and increase the statistical power of this study, the decision was made to include only samples from Caucasian patients. This represents a significant limitation of the study. It is important to acknowledge the existence of racial variation in the incidence and outcome of severe sepsis. Specifically Black and Hispanic populations are associated with higher rates of morbidity and mortality in sepsis, compared to white populations (Barnato et al., 2008, DiMeglio et al., 2018, Linnander et al., 2022). These differences are attributed to both genetic factors, the prevalence of pre-existing comorbidities and inequalities in access to healthcare. For instance, Pneumococcal vaccination reduces the incidence of pneumonia and sepsis, however, minority nursing home patients receive vaccinations at a rate up to 15% lower than white patients placing them at increased risk for recurrent sepsis and adverse long-term outcome (DiMeglio et al., 2018). Interestingly, a single nucleotide polymorphism (SNP) in the *MAL* gene has been shown to confer resistance to septic shock, and is found at a much higher frequency in West Eurasian populations compared to African populations (Ferwerda et al., 2009). By restricting the

study presented in this thesis to patients of European descent, we may have inadvertently limited its generalisability to populations that are disproportionally affected by sepsis.

6.3 Future Directions

6.3.1 Repertoire Reconstruction from Bulk RNA Sequencing

In this thesis, I applied a multiplex PCR reaction to amplify the adaptive immune repertoire for a subset of GAinS patients. However, there is a larger sepsis cohort consisting of approximately 900 patient samples where bulk RNA-seq data is available (Cano-Gamez et al., 2022). This includes patients with CAP and FP sepsis, and a small number of temporal samples. Unfortunately, no healthy controls were sequenced. To validate and expand our findings, while addressing the limitation discussed in **6.2.3**, it would be valuable to perform repertoire reconstruction using the bulk RNA-seq data and a computational tool such TRUST4 or MiXCR, which assemble VDJ sequences from short bulk RNA sequencing reads (Song et al., 2021, Bolotin et al., 2015).

Preliminary analysis using TRUST4 indicates sufficient read depth (>1000 sequences) for the BCR IgH repertoire in a substantial number of sepsis samples, enabling representative repertoire analysis [**Figure 118**]. However, the number of assembled VDJ sequences for TCR α and β chains is low and reconstructed repertoires may not fully represent the overall repertoire. Nonetheless, this approach should identify the abundantly expressed and potentially more clonally expanded TCRs, which are more likely to be involved in antigen binding. TRUST4 yielded minimal TCRγ and δ chains, reflecting their scarcity in peripheral blood, while our in-house multiplex PCR reaction successfully captured an average of 3,095 TCR δ sequences and 3,700 TCR γ sequences in sepsis patients.

Caution must be exercised when comparing the results of our targeted analysis to those obtained using TRUST4, as many assembled VDJ sequences are incomplete due to the short read length. This can impact the accuracy of V gene and allele assignment and estimates of SHM. Additionally, TRUST4 clusters reads with small base pair differences due to SHM or sequencing error into the same contig and generates a representative consensus sequence, which introduces complexities when comparing our in-house pipeline to TRUST4. However, these challenges are outweighed by the benefits of an analysis on a large well-characterised patient cohort.

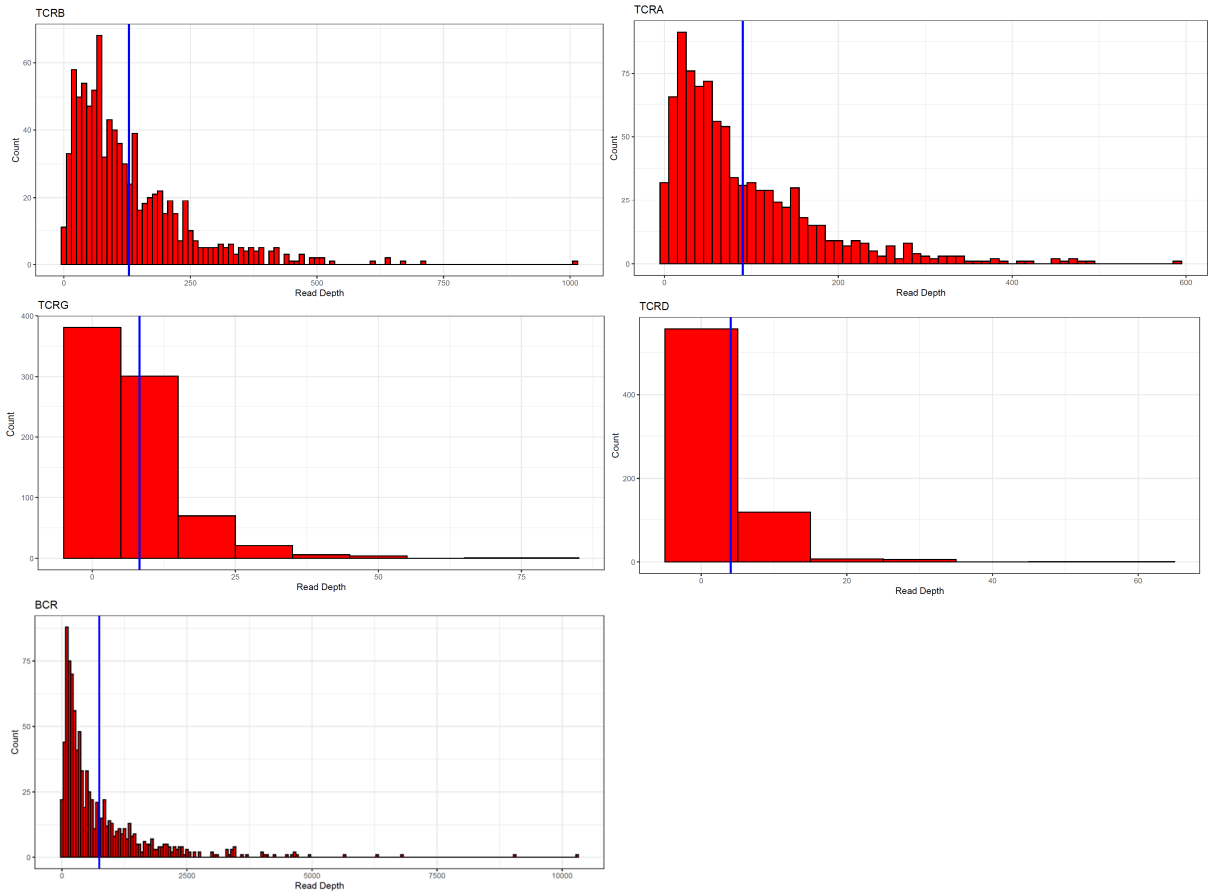


Figure 118. The number of VDJ IgH, TCR α, β, γ and δ sequences identified by TRUST4 for the ~900 sepsis samples for which bulk RNA sequencing data is available.

The mean read depth per chain is shown in blue.

6.3.2 Receptor Specificity

In the future, it would be particularly intriguing to profile TCR and BCR specificity in sepsis patients, to understand the functional implications of disease-specific perturbations to the adaptive immune repertoire, such as loss of diversity and clonal expansion in sepsis.

Matched frozen plasma samples are available for both the sepsis and healthy control patients included in this thesis, making it feasible to use VirScan (CDI Labs) to evaluate the range of viral proteins bound by existing antibodies in these individuals (Shrock et al., 2022). VirScan utilises a phage library consisting of peptides spanning the entire human viral proteome, and immunoprecipitation sequencing (PhIP-Seq) is used to identify the spectrum of targets bound by an individual's plasma.

Similar to sepsis, measles infection is associated with increased morbidity and mortality for years after infection, often due to secondary infections. As with sepsis, the reasons behind this are poorly understood. Mina et al. (2019) used VirScan to screen the antibody targets of 77 unvaccinated children before and 2 months after measles infection. The results show that measles leads to a loss in antibody diversity, measured by the total number of unique pathogen epitopes recognised, while minimal change was observed in the absence of measles infection. The magnitude of effect varied among individuals, with 16% of children experiencing a loss of more than 40% of their repertoire diversity, leading to a state of "immunological amnesia", which may leave them vulnerable to future infections. Repertoire reconstruction occurred after re-exposure on a per-pathogen basis, and it was observed for pathogens such as adenovirus C, influenza A virus, respiratory syncytial virus, human herpesvirus 4 (Epstein-Barr virus; HHV-4), *Streptococcus pneumoniae* and *Staphylococcus aureu*s. A similar study could be performed with acute sepsis patients and convalescent samples to determine whether immunological anaemia occurs, due to lymphopoiesis and loss

of diversity. This would be essential to link perturbations to the adaptive immune repertoire with repeated infection and mortality in post-sepsis syndrome.

The CDI Lab also offer the same technology for screening plasma/serum against the full NCBI human proteome (HuScan) in order to identify autoantibodies. This is also an interesting avenue for further research in sepsis. Previous studies have indicated that 46% of severe sepsis patients have at least one statistically significantly elevated autoantibody compared to controls. This includes autoantibodies against the lung-specific potassium channel regulator, gastric ATPase and notably, high levels against certain pro-inflammatory cytokines (interleukin-6, INFγ). The level of these autoantibodies is initially low, but increases to a peak 7-14 days after sepsis-onset which might reflect autoimmunisation (Burbelo et al., 2010). Similar findings have been reported in life-threatening COVID-19 pneumonia, where at least 10% of patients carried neutralising antibodies against IFN-α2 and/or –ω which were not found in any of the asymptomatic individuals tested. Interestingly, the presence of autoantibodies was biased towards male sex and most individuals were over the age of 65, which aligns with the demographics commonly associated with sepsis. A review human autoantibodies in infectious disease is provided by Puel et al. (2022).

Determining TCR specificity is more challenging, because of the requirement for recognition via peptide-MHC complexed antigen presentation. However, an increasing number of computational tools have been developed in order to predict TCR antigen specificity, which could be applied to this dataset (Fischer et al., 2020). For instance, TCRMatch which predicts receptor specificity based on sequence similarity to previously characterised receptors (Chronister et al., 2021) .

6.3.3 Integration of Multi-Omic Data

To gain insights into the regulation of the adaptive immune repertoire and the downstream effects of repertoire alterations, in **chapters 4** and **5** of this thesis, BCR and TCR repertoire modules were integrated with bulk RNA-seq gene expression data. The success of this innovative analysis serves as a basis for future research, involving the integration of additional experimental modalities. As part of the GAinS project multiple genomic and functional datasets have been generated including RNA-seq and microarray transcriptomics, exome sequencing (n>500) and metagenomics (n>500) which each capture unique but related aspects of the host sepsis response.

For a subset of the patients included in this thesis, plasma proteomic data was generated using TimsTOF mass spectrometry as part of another project. It would be valuable to investigate how changes in the BCR repertoire, such as immunoglobulin isotype usage, correlate with the composition of the immunoglobulin proteome (Mi et al., 2022). This is a particularly important analysis, due to the differential contribution of various B cell types to the proteome compared to the adaptive immune repertoire. The plasma immunoglobulin proteome is influenced by antibody-secreting B cell populations (plasma cells and plasmablasts), whereas BCR repertoire sequencing captures RNA sequences that give rise to both BCRs and antibodies. The majority of circulating B cells are naive B cells (60-70%) - antibody secreting cells comprise a much smaller proportion (~2%). However, their RNA is over-represented due to increased antibody RNA synthesis which can skew the repertoire. Thus, findings in BCR repertoire studies do not always translate to the proteome.

The general consensus in the literature is the presence of subnormal-levels of plasma immunoglobulins among sepsis patients at diagnosis, including IgM, G and A (Nierhaus et al., 2020). However, IgA1, A2 and G1 gene expression was elevated in sepsis compared to

health and expression of all immunoglobulin isotypes increased significantly during disease course. IgM expression was significantly lower than health, but also increased significantly during disease course. Although the mechanisms underpinning the low-levels of plasma immunoglobulin are poorly understood, researchers have suggested it may either result from reduced production and secretion, their redistribution into inflamed tissues or over-utilisation by the complement system and vascular leakage. Given the elevated gene expression, which increased during disease, our results suggest that reduced production is not the case, although dysfunctional translation or secretion cannot be ruled out (Nierhaus et al., 2020). Therefore a more detailed analysis comparing the BCR repertoire to the available proteomic data for these patients is warranted.

It will also be valuable to explore the influence of host immunogenic variation on various aspects of the adaptive immune repertoire, such as TCR diversity and V gene usage, as well as its impact on sepsis outcomes. This can be achieved by performing human leukocyte antigen (HLA) typing on sepsis patients and several computational approaches have been developed to use bulk RNA-seq data. Sequencing reads are mapped against a reference database of HLA alleles, from which HLA type and locus-specific expression levels can be determined (Boegel et al., 2012). Previous research has demonstrated a positive correlation between TCR diversity and polymorphisms at the human leukocyte class I antigen (Krishna et al., 2020). Genetic variation in MHC proteins is also known to shape TCRV gene usage (Sharon et al., 2016). As biased V gene usage was the major finding in the sepsis $\alpha\beta$ TCR repertoire, understanding immunogenic variation may help to ascertain whether this observation precedes the development of sepsis, is antigen driven or a mix of both.

6.4 Conclusion

The overarching aim of this thesis was to characterise the BCR and TCR repertoire during acute sepsis using multiplex PCR. In doing so I have generated a substantial repertoire dataset of ~200 patient samples from 72 individuals, which will be a valuable resource for future integrative sepsis research. The analysis of such a large complex dataset required the development of an innovative dimensional reduction analysis pipeline (rWGCNA) to understand regulation of the adaptive immune repertoire, which can be used by the wider repertoire community. Using this approach, I have uncovered novel sepsis biology associated with mortality, potentially offering opportunities for the development of new therapeutics.

A Appendices: Materials and Methods

Comorbidity	Number of Patients	Percentage of Patients (%) (n=72)
Myocardial Infarction	10	13.89
Congestive Heart Failure	1	1.39
Peripheral Vascular Disease	1	1.39
Cerebrovascular Accident (CVA)	5	6.94
Chronic Obstructive Pulmonary Disease (COPD)	20	27.78
Connective Tissue Rheumatoid Arthritis (RA)	2	2.78
Peptic Ulcer Disease (PUD)	1	1.39
Liver Disease	2	2.78
Diabetes	11	15.28
Chronic Kidney Disease	7	9.72
Tumour	4	5.56
Metastatic Tumour	1	1.39
Leukaemia	1	1.39
	Mean	**Standard Deviation**
Charlson Index	4.04	1.86

Table A. 1. Significant patient comorbidities for sepsis cohorts 1 and 2, as recorded in the eCRF.

The mean and standard deviation for the Charlson Index is also reported. Results are reported to 2 decimal places.

#ID	Age (years)	Sex	Day 1	Day 2	Day 3	Day 4	Day 5
HV_105	34	F	X		X	X	
HV_106	70	F	X		X		X
HV_107	52	F	X		X		X
HV_108	51	M	X		X		X
HV_109	52	M	X		X		X
HV_110	56	F	X		X		X
HV_111	61	F	X		X	X	
HV_112	35	M	X	X		X	
HV_113	53	M	X	X		X	
HV_114	60	M	X		X	X	
HV_115	55	M	X	X	X		
HV_116	65	F	X		X	X	

Table A. 2. Healthy volunteer recruitment timeline.

'X' denotes day of sampling. Blood was drawn with a minimum of 24 hours between sampling.

Sex	Number of Healthy Volunteers	Percentage of Healthy Volunteers (%) (n=12)	Mean Age	Standard Deviation Age	Min Age	Max Age	% Smokers
Male	6	50	51.00	8.46	35	60	0.00
Female	6	50	56.33	12.66	34	70	16.67

Table A. 3. Healthy volunteer demographics including age, sex and smoking status.

Name	Type	Sequence
VH1-FR1_F1	F	AAGCAGTGGTATCAACGCAGAGTCGCAGGGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F2	F	AAGCAGTGGTATCAACGCAGAGCTCTGCAGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F3	F	AAGCAGTGGTATCAACGCAGAGCCTAGGTGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F4	F	AAGCAGTGGTATCAACGCAGAGGGATCAAGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F5	F	AAGCAGTGGTATCAACGCAGAGGCAAGATGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F6	F	AAGCAGTGGTATCAACGCAGAGATGGAGAGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F7	F	AAGCAGTGGTATCAACGCAGAGCTCGATGGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F8	F	AAGCAGTGGTATCAACGCAGAGGCTCGAAGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F9	F	AAGCAGTGGTATCAACGCAGAGACCAACTGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F10	F	AAGCAGTGGTATCAACGCAGAGCCGGTACGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F11	F	AAGCAGTGGTATCAACGCAGAGAACTCCGGGCCTCAGTGAAGGTCTCCTGCAAG
VH1-FR1_F12	F	AAGCAGTGGTATCAACGCAGAGTTGAAGTGGCCTCAGTGAAGGTCTCCTGCAAG
VH2-FR1_F1	F	AAGCAGTGGTATCAACGCAGAGTCGCAGGGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F2	F	AAGCAGTGGTATCAACGCAGAGCTCTGCAGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F3	F	AAGCAGTGGTATCAACGCAGAGCCTAGGTGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F4	F	AAGCAGTGGTATCAACGCAGAGGGATCAAGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F5	F	AAGCAGTGGTATCAACGCAGAGGCAAGATGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F6	F	AAGCAGTGGTATCAACGCAGAGATGGAGAGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F7	F	AAGCAGTGGTATCAACGCAGAGCTCGATGGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F8	F	AAGCAGTGGTATCAACGCAGAGGCTCGAAGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F9	F	AAGCAGTGGTATCAACGCAGAGACCAACTGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F10	F	AAGCAGTGGTATCAACGCAGAGCCGGTACGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F11	F	AAGCAGTGGTATCAACGCAGAGAACTCCGGTCTGGTCCTACGCTGGTGAAACCC
VH2-FR1_F12	F	AAGCAGTGGTATCAACGCAGAGTTGAAGTGTCTGGTCCTACGCTGGTGAAACCC
VH3-FR1_F1	F	AAGCAGTGGTATCAACGCAGAGTCGCAGGCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F2	F	AAGCAGTGGTATCAACGCAGAGCTCTGCACTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F3	F	AAGCAGTGGTATCAACGCAGAGCCTAGGTCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F4	F	AAGCAGTGGTATCAACGCAGAGGGATCAACTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F5	F	AAGCAGTGGTATCAACGCAGAGGCAAGATCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F6	F	AAGCAGTGGTATCAACGCAGAGATGGAGACTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F7	F	AAGCAGTGGTATCAACGCAGAGCTCGATGCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F8	F	AAGCAGTGGTATCAACGCAGAGGCTCGAACTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F9	F	AAGCAGTGGTATCAACGCAGAGACCAACTCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F10	F	AAGCAGTGGTATCAACGCAGAGCCGGTACCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F11	F	AAGCAGTGGTATCAACGCAGAGAACTCCGCTGGGGGGGTCCCTGAGACTCTCCTG
VH3-FR1_F12	F	AAGCAGTGGTATCAACGCAGAGTTGAAGTCTGGGGGGGTCCCTGAGACTCTCCTG
VH4-FR1_F1	F	AAGCAGTGGTATCAACGCAGAGTCGCAGGCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F2	F	AAGCAGTGGTATCAACGCAGAGCTCTGCACTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F3	F	AAGCAGTGGTATCAACGCAGAGCCTAGGTCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F4	F	AAGCAGTGGTATCAACGCAGAGGGATCAACTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F5	F	AAGCAGTGGTATCAACGCAGAGGCAAGATCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F6	F	AAGCAGTGGTATCAACGCAGAGATGGAGACTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F7	F	AAGCAGTGGTATCAACGCAGAGCTCGATGCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F8	F	AAGCAGTGGTATCAACGCAGAGGCTCGAACTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F9	F	AAGCAGTGGTATCAACGCAGAGACCAACTCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F10	F	AAGCAGTGGTATCAACGCAGAGCCGGTACCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F11	F	AAGCAGTGGTATCAACGCAGAGAACTCCGCTTCGGAGACCCTGTCCCTCACCTG
VH4-FR1_F12	F	AAGCAGTGGTATCAACGCAGAGTTGAAGTCTTCGGAGACCCTGTCCCTCACCTG
VH5-FR1_F1	F	AAGCAGTGGTATCAACGCAGAGTCGCAGGCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F2	F	AAGCAGTGGTATCAACGCAGAGCTCTGCACGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F3	F	AAGCAGTGGTATCAACGCAGAGCCTAGGTCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F4	F	AAGCAGTGGTATCAACGCAGAGGGATCAACGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F5	F	AAGCAGTGGTATCAACGCAGAGGCAAGATCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F6	F	AAGCAGTGGTATCAACGCAGAGATGGAGACGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F7	F	AAGCAGTGGTATCAACGCAGAGCTCGATGCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F8	F	AAGCAGTGGTATCAACGCAGAGGCTCGAACGGGGAGTCTCTGAAGATCTCCTGT

Name	Type	Sequence
VH5-FR1_F9	F	AAGCAGTGGTATCAACGCAGAGACCAACTCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F10	F	AAGCAGTGGTATCAACGCAGAGCCGGTACCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F11	F	AAGCAGTGGTATCAACGCAGAGAACTCCGCGGGGAGTCTCTGAAGATCTCCTGT
VH5-FR1_F12	F	AAGCAGTGGTATCAACGCAGAGTTGAAGTCGGGGAGTCTCTGAAGATCTCCTGT
VH6-FR1_F1	F	AAGCAGTGGTATCAACGCAGAGTCGCAGGTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F2	F	AAGCAGTGGTATCAACGCAGAGCTCTGCATCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F3	F	AAGCAGTGGTATCAACGCAGAGCCTAGGTTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F4	F	AAGCAGTGGTATCAACGCAGAGGGATCAATCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F5	F	AAGCAGTGGTATCAACGCAGAGGCAAGATTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F6	F	AAGCAGTGGTATCAACGCAGAGATGGAGATCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F7	F	AAGCAGTGGTATCAACGCAGAGCTCGATGTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F8	F	AAGCAGTGGTATCAACGCAGAGGCTCGAATCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F9	F	AAGCAGTGGTATCAACGCAGAGACCAACTTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F10	F	AAGCAGTGGTATCAACGCAGAGCCGGTACTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F11	F	AAGCAGTGGTATCAACGCAGAGAACTCCGTCGCAGACCCTCTCACTCACCTGTG
VH6-FR1_F12	F	AAGCAGTGGTATCAACGCAGAGTTGAAGTTCGCAGACCCTCTCACTCACCTGTG
CNUS	R	TGTCCAGCACGCTTCAGGCT
IGHA constant	RT	TGTCCAGCACGCTTCAGGCTNNNNTNNNNTNNNNGAYGACCACGTTCCCATCT
IGHM constant	RT	TGTCCAGCACGCTTCAGGCTNNNNTNNNNTNNNNTCGTATCCGACGGGGAATTC
IGHD constant	RT	TGTCCAGCACGCTTCAGGCTNNNNTNNNNTNNNNGGGCTGTTATCCTTTGGGTG
IGHE constant	RT	TGTCCAGCACGCTTCAGGCTNNNNTNNNNTNNNNAGAGTCACGGAGGTGGCATT
IGHG constant	RT	TGTCCAGCACGCTTCAGGCTNNNNTNNNNTNNNNAGTAGTCCTTGACCAGGCAG
IGHA constant	RT	TGTCCAGCACGCTTCAGGCTNNNNTNNNNTNNNNGAYGACCACGTTCCCATCT

Table A. 4. BCR primers sequences.

Included are the reverse transcription primers (RT), PCR forward primers (F) and PCR reverse primers (R). The PCR forward primer names are coloured according to pooled primer mix *e.g.* VH1/2/3/4/5/6_F1 are pooled together into one primer mix called barcode 1 (BC1).

TCR	Name	Type	Sequence
α	TRAC_RT	RT	CAGCACGCTTCAGGCTNNNNTNNNNTNNNNTGGCAGACAGACTTGTCACTGG
β	TRBC_RT	RT	CAGCACGCTTCAGGCTNNNNTNNNNTNNNNAGATCTCTGCTTCTGATGGCTC
γ	TRGC_RT_Ttag	RT	TCGGCATTCCTGCTGAACCGCTCTTCCGATCTTNNNNTNNNNTNNNNGAAGGAAGAAAAA TAGTGGGCT
δ	TRDC_RT_1_T tag	RT	TCGGCATTCCTGCTGAACCGCTCTTCCGATCTTNNNNTNNNNTNNNNCGGATGGTTTGGT ATGAGGCT
α	TRAV1-2	F	GGACAAARCMTTGASCA
α	TRAV2	F	AAGGACCAAGTGTTTCA
α	TRAV3	F	GCTCAGTCAGTGGCTCA
α	TRAV4	F	CTTGCTAAGACCACCCA
α	TRAV5	F	GGAGAGGATGTGGAGCAGA
α	TRAV6	F	AGCCAAAAGATAGAACAGA
α	TRAV7	F	GAAAACCAGGTGGAGCACA
α	TRAV8B	F	GCCCAGTCKGTGASCCAGCW
α	TRAV9	F	GGARATTCAGTGRYCCAGA
α	TRAV10	F	TGGAGCAGAGTCCTCAGT
α	TRAV12	F	ACCMYTCARTGTTCCAGA
α	TRAV13	F	GAGAGARTGTGGRGCWGCA
α	TRAV14	F	CTCAAACCCAACCAGGAAT
α	TRAV16	F	AGAGAGTGACTCAGCCCGA
α	TRAV17	F	AGTCAACAGGGAGAAGAGGA
α	TRAV18	F	GGAGACTCGGTTACCCAGA
α	TRAV19	F	GCTCAGAAGGTAACTCAA
α	TRAV20	F	GAAGACCAGGTGACGCAGA
α	TRAV21	F	AAACAGGAGGTGACRCAGA
α	TRAV22	F	GGAATACAAGTGGAGCAGA
α	TRAV23	F	CAGCAGGTGAAACAAAGT
α	TRAV24	F	TACTGAACGTGGAACAARG
α	TRAV25	F	GGACAACAGGTAATGCAAA
α	TRAV26	F	GATGCTAAGACCACMCAGC
α	TRAV27	F	ACCCAGCTGCTGGAGCAGA
α	TRAV29	F	GACCAGCAAGTTAAGC
α	TRAV30	F	CAACAACCAGTGCAGAGTC
α	TRAV34	F	AGCCAAGAACTGGAGCAGA
α	TRAV35	F	GGTCAACAGCTGAATCAGA
α	TRAV36	F	GAAGACAAGGTGGTACAAA
α	TRAV38	F	GCYCAGACAGTCACTCAGT
α	TRAV39	F	GAAAGTGGAACAAAACCCT
α	TRAV40	F	AGCAATTCAGTCAAGCAGA
α	TRAV41	F	AAGTGGAGCAGAGTCCTCA
β	TRBV2	F	GAACCTGAAGTCACCCAGA
β	TRBV3	F	GACACAGCCGTTTCCCAGA
β	TRBV4	F	ACKGRAGTTACSCAGACA
β	TRBV5	F	RCTGGAGTCACHCAAAST
β	TRBV6	F	TGCTGGKRTCACYCAG
β	TRBV7	F	WDCTGGAGTYTCCCA
β	TRBV9	F	GATTCTGGAGTCACACAAA
β	TRBV10	F	GATGCTGRAATCACCCAGA
β	TRBV11	F	GAAGCTGRAGTGGYYCAGT
β	TRBV12	F	GATGCTRGAGTYAYCCAG
β	TRBV13B	F	GCTGCTGGAGTCATCCAGT
β	TRBV14	F	GAAGCTGGAGTTACTCAGT
β	TRBV15	F	GATGCCATGGTCATCCAGA
β	TRBV16	F	GGTGAAGAAGTCGCCCAGA
β	TRBV19	F	GATGGTGGAATCACTCAGT
β	TRBV18B	F	AATGCCGGCGTCATGCAGA
β	TRBV20	F	GGTGCTGTCGTCTCTCAAC
β	TRBV24	F	GATGCTGATGTTACCCAGA
β	TRBV25	F	GAAGCTGACATCTACCAGA
β	TRBV27	F	GAAGCCCAAGTGACCCAGA
β	TRBV28	F	GATGTGAAAGTAACCCAGA
β	TRBV29	F	AGTGCTGTCATCTCTC
β	TRBV30	F	TCTCAGACTATTCATCAAT
γ	TRGV_1	F	CCAACTTGGAAGGGRGAAC
γ	TRGV_2	F	ACCTAGAGCAACCTCAAATTTCC

γ	TRGV_3	F	GGAGCAGTTCCAGCTATCCA
γ	TRGV_4	F	TTCCAGACCAGCAAATAAGAGTG
δ	TRDV1	F	AGAAGGTTACTCAAGCCCAGTC
δ	TRDV1	F	ATTGAGTTGGTGCCTGAACAC
δ	TRDV3	F	TGACAAAGTAACCCAGAGTTCCC
δ	TRAV29 (DV5)	F	GACCAGCAAGTTAAGC
δ	TRAV14 (DV4)	F	CTCAAACCCAACCAGGAAT
δ	TRAV23 (DV6)	F	CAGCAGGTGAAACAAAGT
δ	TRAV36 (DV7)	F	GAAGACAAGGTGGTACAA
δ	TRAV38 (DV8)	F	GCYCAGACAGTCACTCAGT
α/β	CNUS2_BC1	R	CGCTCTTCCGATCTTaggtattCAGCACGCTTCAGGCT
α/β	CNUS2_BC2	R	CGCTCTTCCGATCTTattaaggCAGCACGCTTCAGGCT
α/β	CNUS2_BC3	R	CGCTCTTCCGATCTTtaattctCAGCACGCTTCAGGCT
α/β	CNUS2_BC4	R	CGCTCTTCCGATCTTataagctCAGCACGCTTCAGGCT
α/β	CNUS2_BC5	R	CGCTCTTCCGATCTTttatgatCAGCACGCTTCAGGCT
α/β	CNUS2_BC6	R	CGCTCTTCCGATCTTgagttaaCAGCACGCTTCAGGCT
α/β	CNUS2_BC7	R	CGCTCTTCCGATCTTattcttgCAGCACGCTTCAGGCT
α/β	CNUS2_BC8	R	CGCTCTTCCGATCTTttagcttCAGCACGCTTCAGGCT
α/β	CNUS2_BC9	R	CGCTCTTCCGATCTTatcttgcCAGCACGCTTCAGGCT
α/β	CNUS2_BC10	R	CGCTCTTCCGATCTTtctccatCAGCACGCTTCAGGCT
α/β	CNUS2_BC11	R	CGCTCTTCCGATCTTacctaggCAGCACGCTTCAGGCT
α/β	CNUS2_BC12	R	CGCTCTTCCGATCTTttgatccCAGCACGCTTCAGGCT
γ/δ	TCR_BC1_DG	R	CGAGATaggtattTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC2_DG	R	CGAGATattaaggTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC3_DG	R	CGAGATtaattctTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC4_DG	R	CGAGATataagctTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC5_DG	R	CGAGATttatgatTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC6_DG	R	CGAGATgagttaaTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC7_DG	R	CGAGATattcttgTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC8_DG	R	CGAGATttagcttTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC9_DG	R	CGAGATcctatctTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC10_DG	R	CGAGATtctccatTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC11_DG	R	CGAGATacctaggTCGGCATTCCTGCTGAACC
γ/δ	TCR_BC12_DG	R	CGAGATgtaattgTCGGCATTCCTGCTGAACC

Table A. 5. TCR primer sequences.

Included are the TCR reverse transcription primers (RT), PCR forward primers (F) and PCR reverse primers (R).

Region	IMGT Annotation	Definition
V gene	Functional **(F)**	Open read frame without stop codon, defect in splicing sites, recombination signals or regulatory elements.
V gene	Pseudogene **(P)**	■ Stop codon or frameshift mutations. ■ Part of the V-region is missing (too mutated to be described).
V gene	Open Read Frame **(ORF)**	■ Alterations have been described in the splicing sites, recombination signals or regulatory elements. ■ Changes to conserved amino acids may lead to incorrect folding. ■ Entity is an orphan.
VDJ	Productive	■ The coding region has an open reading frame with no stop codon and no defect described in initiation codon, splicing sites or regulatory elements and an in-frame junction.
VDJ	Unproductive	■ Contains a stop codon or frameshift mutations. ■ Mutations affecting the initiation codon. ■ Defects in the splicing sites and/or regulatory elements. ■ Changes to conserved amino acids demonstrated as leading to incorrect folding.

Table A. 6. IMGT VDJ and V gene annotations.

Definitions sourced from:
www.imgt.org/IMGTScientificChart/SequenceDescription/IMGTfunctionality.html

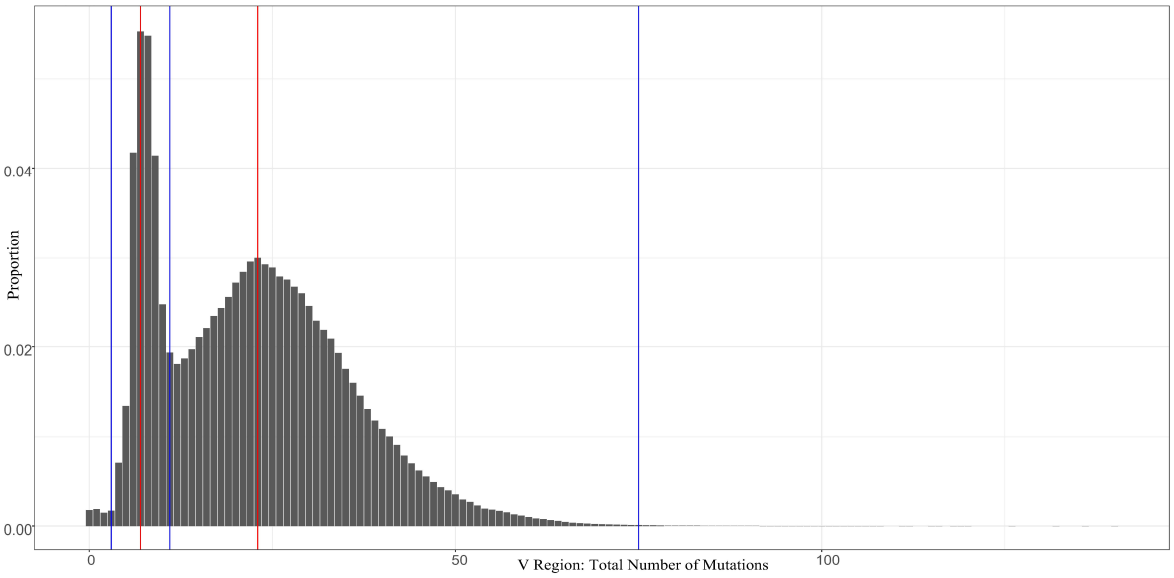


Figure A. 1. BCR SHM distribution for all productive unique VDJ sequences.

The histogram clearly shows a bimodal distribution. Thresholds used to classify sequences according to SHM burden are shown in blue, while red lines denote local maximum.

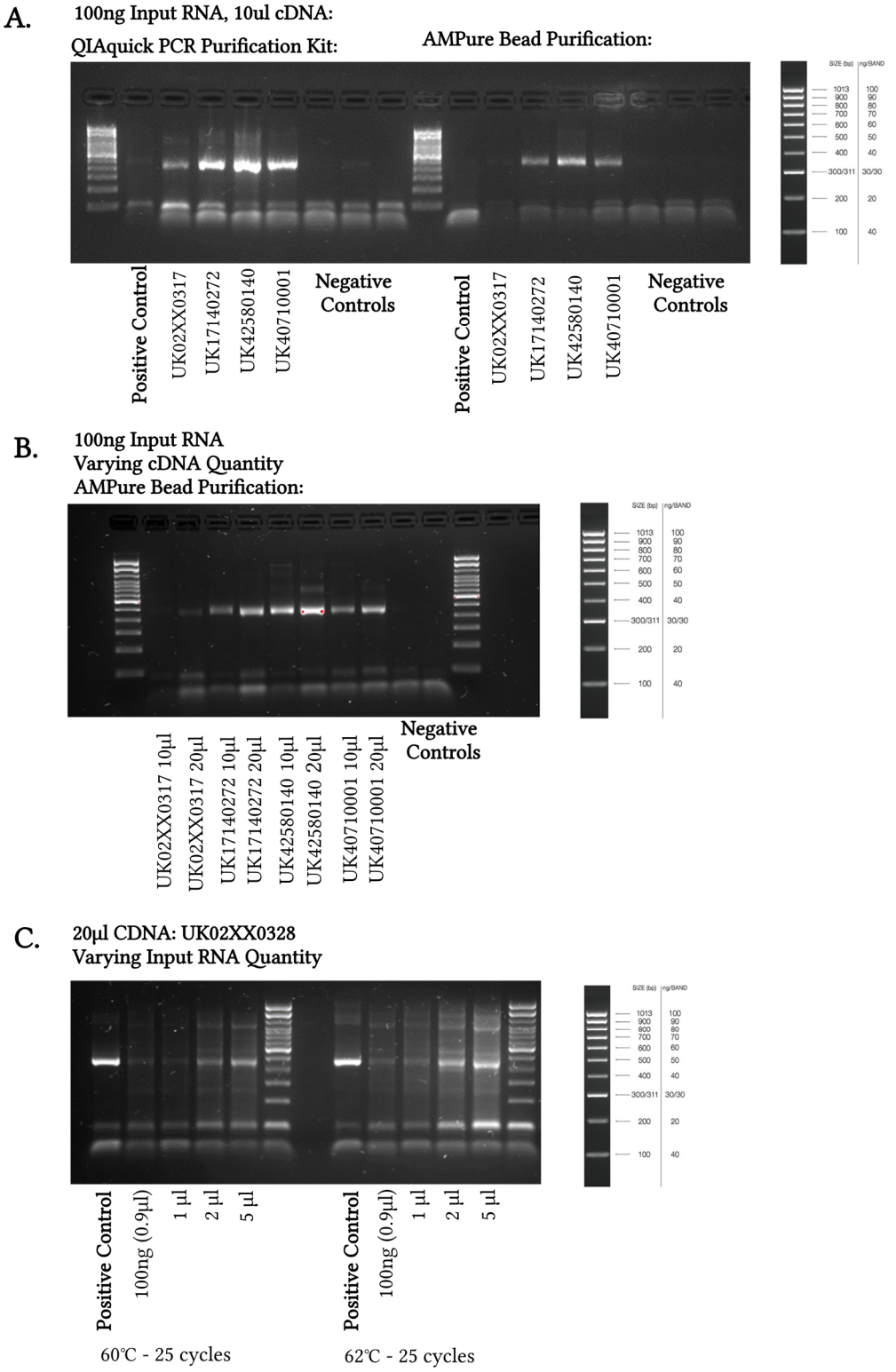


Figure A. 2. Agarose gels demonstrating a subset of BCR repertoire amplification protocol optimisations.

A) Post-PCR amplicons: a comparison between bead-based and column-based RNA extraction methods. **B)** Post-PCR amplicons: comparing different cDNA input quantities for PCR. **C)** Post-PCR amplicons: comparison of varying input RNA quantities. The Bioline ladder reference is displayed on the right side of the gel images.

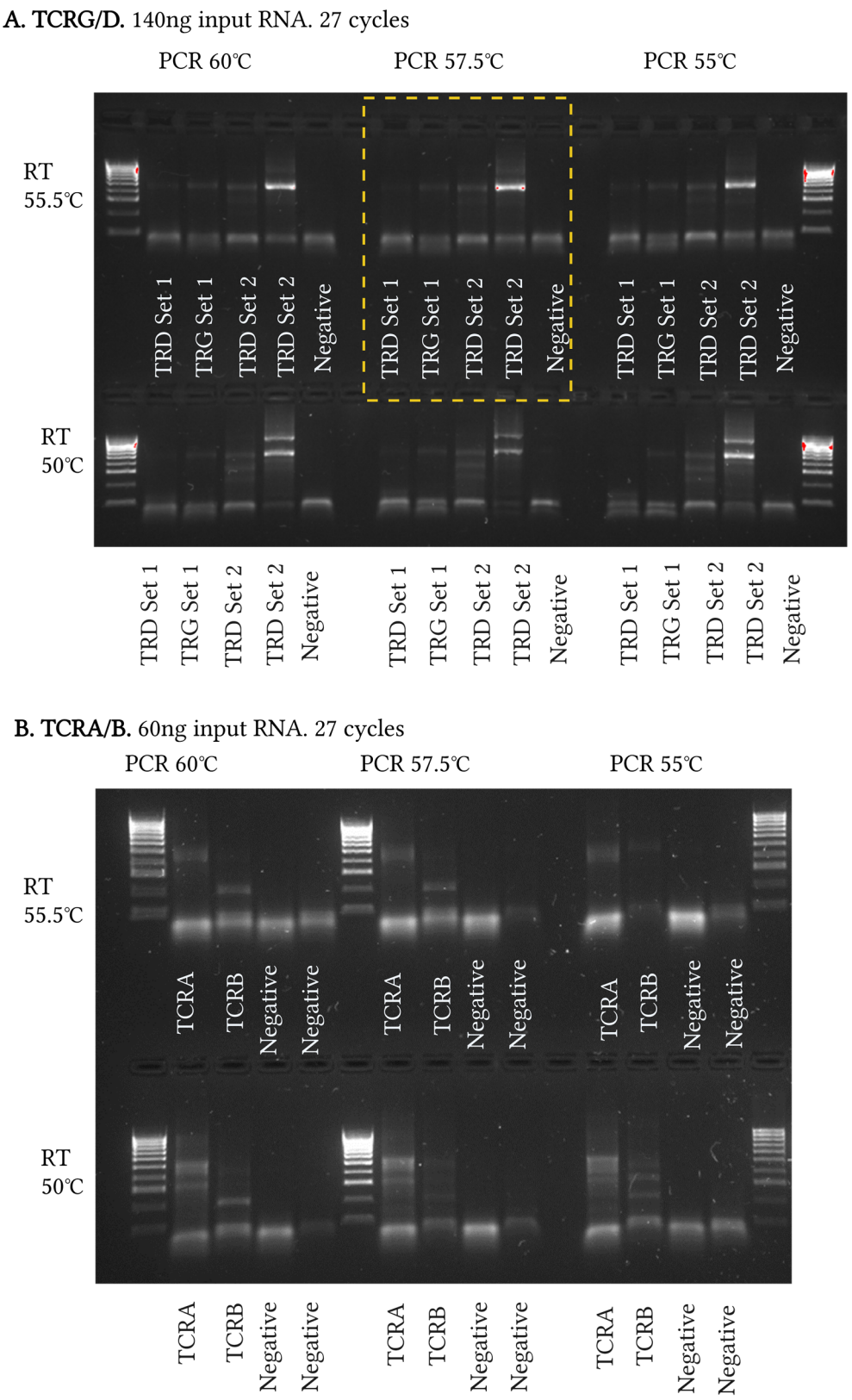


Figure A. 3. Agarose gels demonstrating a subset of TCR repertoire amplification protocol optimisations.

RT and PCR temperatures were optimised using gradient PCR for: **A)** TCRG and TCRD, also comparing two different primer sets (1 and 2) and **B)** TCRA and TCRB.

B Appendices: The Development of Computational Methods to Analyse the BCR and TCR Repertoire

Number of Samples	BCR Libraries ID	Project ID	Sequencing Lane ID	Cohort	Date
81	1-11	P210054	1	Sepsis Cohort 1	02/03/2021
51	12-18	P210054	2	Sepsis Cohort 1	29/03/2021
23	19-21	P210425	9	Sepsis Cohort 2	10/11/2021
22	22-24	P210425	10	Sepsis Cohort 2	18/11/2021
16	25,26,28	P220161	11	Healthy Volunteers	13/05/2022
22	27,29,30	P220161	12	Healthy Volunteers	13/05/2022

Table B. 1. Sequencing timeline for the 30 BCR IgH repertoire libraries generated using multiplex PCR.

Due to recruitment timelines amplification and sequencing of healthy volunteer and sepsis samples was performed separately. Sequencing was performed by the Oxford Genomics Centre (OGC).

	Percentage of Repertoire		
Isotype	**Mean**	**SD**	**NSD**
IGHA1*	16.36	1.40	0.09
IGHA2	4.92	0.57	0.12
IGHD	7.83	2.01	0.26
IGHE	0.07	0.02	0.31
IGHEP2	0.05	0.02	0.51
IGHG1	6.17	1.18	0.19
IGHG2*	10.00	1.61	0.16
IGHG3	0.90	0.28	0.31
IGHG4	0.12	0.06	0.50
IGHGP	0.03	0.02	0.60
IGHM*	53.55	4.33	0.08

Table B. 2. QC summary statistics: isotype usage for the 10 BCR technical replicates.

Statistics include the mean, standard deviation (SD) and normalised standard deviation (NSD=SD/Mean) for the percentage of the repertoire comprised of each class of antibody isotype. Estimated usage for the most common isotypes detected in the blood (*) was more reproducible across technical replicates than for rare isotypes. Results are reported to 2 decimal places.

	Percentage of Unique VDJ Reads		
Gene	**Mean**	**SD**	**NSD**
IGHJ1	1.97	0.18	0.09
IGHJ2	3.99	0.44	0.11
IGHJ3	12.09	0.75	0.06
IGHJ4	50.71	2.22	0.04
IGHJ5	11.82	0.43	0.04
IGHJ6	19.42	1.34	0.07
IGHV1	34.00	7.86	0.23
IGHV2	4.36	1.56	0.36
IGHV3	36.91	5.73	0.16
IGHV4	16.91	2.93	0.17
IGHV5	6.60	0.77	0.12
IGHV6	1.12	0.20	0.18
IGHV7	0.08	0.05	0.62

Table B. 3. QC summary statistics: BCR V/J gene family usage for the 10 technical replicates included across sequencing lanes.

Statistics including the mean, standard deviation (SD) and normalised standard deviation (NSD=SD/Mean) for the percentage of the repertoire comprised of each V gene family/J gene. Results are reported to 2 decimal places.

Method	R Package	Summary
Mean, Median and Random	Hmisc	Missing data was imputed using the function *hmisc::impute()*. The 'function' argument was set to either mean, median or random. For mean and median imputation, missing data was imputed using the column-wise mean/median. By comparison, for random imputation, values were drawn at random to replace missing data and if there were multiple missing values in one column, imputed values were not forced to be the same. These methods do not consider correlation between explanatory variables.
Regularised PCA and Expectation Maximisation (EM) PCA	MissMDA	Both regularised PCA and EM-PCA (expectation-maximisation) use an iterative PCA algorithm to impute missing data (Josse and Husson, 2016). The first step uses a random but reasonable value drawn from a Gaussian distribution with mean and median determined from the observed values, to replace missing data. PCA is performed on the completed data and the value fitted by the PCA is used to predict a new value for the original missing data. The procedure is repeated using the new value until convergence. The basic method is known as the iterative EM-PCA algorithm. However, over-fitting is very common when there are many missing values, therefore a regularised version of the PCA algorithm can be used instead.
k-nearest-neighbour (kNN)	VIM	First a distance matrix which is NA aware is generated using an extension of the Gower distance, where the distance between two observations is the weighted mean of the contributions of each variable, and the weight is the importance of the variable. Then aggregation of the k values of the nearest neighbours for NA values is used to impute the missing data. For continuous variables the median is used by default (Kowarik and Templ, 2016).
Random Forest	missForest	Missing values are first filled using an initial guess, which under default settings is the mean (but it can be a prediction from another model). The dataset is divided into a training set and a prediction set according to whether a value is missing in the original data. A RF is fitted on observed values using response $y^{(s)}_{obs}$ and predictors $x^{(s)}_{obs}$ (training set). The missing response values $y^{(s)}_{mis}$ are imputed by applying the trained RF to the prediction set. The imputation procedure is repeated and the model decides whether to adjust predictions or keep them the same, until a stopping criteria is met or a maximum number of user-specified iterations are performed (Stekhoven and Buhlmann, 2012).
Additive Regression	Hmisc	Additive regression imputation was performed using the R package 'Hmisc' using the *aregImpute()* function. A flexible additive model is fitted on a subsample of the original data where the variable being imputed is not missing. The fitted flexible model is then used to predict the target value for all of the original observations including the missing value. This process is repeated using different bootstrap resamples for each of the multiple imputations. Then, predictive mean matching is used to impute the missing values, by using all donors with predicted values closest to the predicted value for the missing value. One of these close cases is drawn from a multinomial distribution with probabilities derived from distance weights and the missing value is replaced by the observed value of this close case. Predictive mean matching helps generate plausible values when imputing missing data.

Table B. 4. Missing data imputation methods including associated R packages.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_{imp} - x_{obs})^2}{N}}$$

Equation B. 1. Root mean squared error (RMSE).

X_{imp} is the imputed value and X_{obs} is the original value.

$$Normalised\ RMSE = \frac{RMSE}{(y_{max} - y_{\min})}$$

Equation B. 2. Normalised RMSE.

	NRMSE: Imputation Method							
NA Seed Frequency	**Mean**	**Median**	**Random**	**missMDA1**	**missMDA2**	**kNN**	**MissForest**	**Additive Regression**
0.05	0.36	0.37	0.53	0.29	0.33	0.30	0.20	0.78
0.10	0.39	0.40	0.52	0.27	0.31	0.29	0.22	0.59
0.15	0.40	0.41	0.53	0.29	0.32	0.34	0.24	0.60
0.20	0.39	0.40	0.56	0.28	0.33	0.34	0.24	0.66
0.25	0.38	0.39	0.58	0.29	0.35	0.34	0.24	0.62
0.30	0.39	0.40	0.56	0.29	0.38	0.36	0.26	0.60
0.35	0.42	0.43	0.58	0.33	0.38	0.40	0.31	0.62
0.40	0.38	0.39	0.56	0.29	0.34	0.36	0.26	0.61
0.45	0.41	0.42	0.55	0.33	0.38	0.39	0.30	0.61
0.50	0.39	0.40	0.55	0.29	0.37	0.38	0.27	0.54

Table B. 5. NRMSE for each missing data imputation method for feature-wise NA seeding applied to the complete-cases BCR repertoire feature matrix.

Results are reported to 2 decimal places.

$$Distance = \frac{1 - cor(x')}{2}$$

Equation B. 3. Converting the repertoire feature correlation matrix into a distance matrix.

Anti-correlated features were treated as more distant than correlated features.

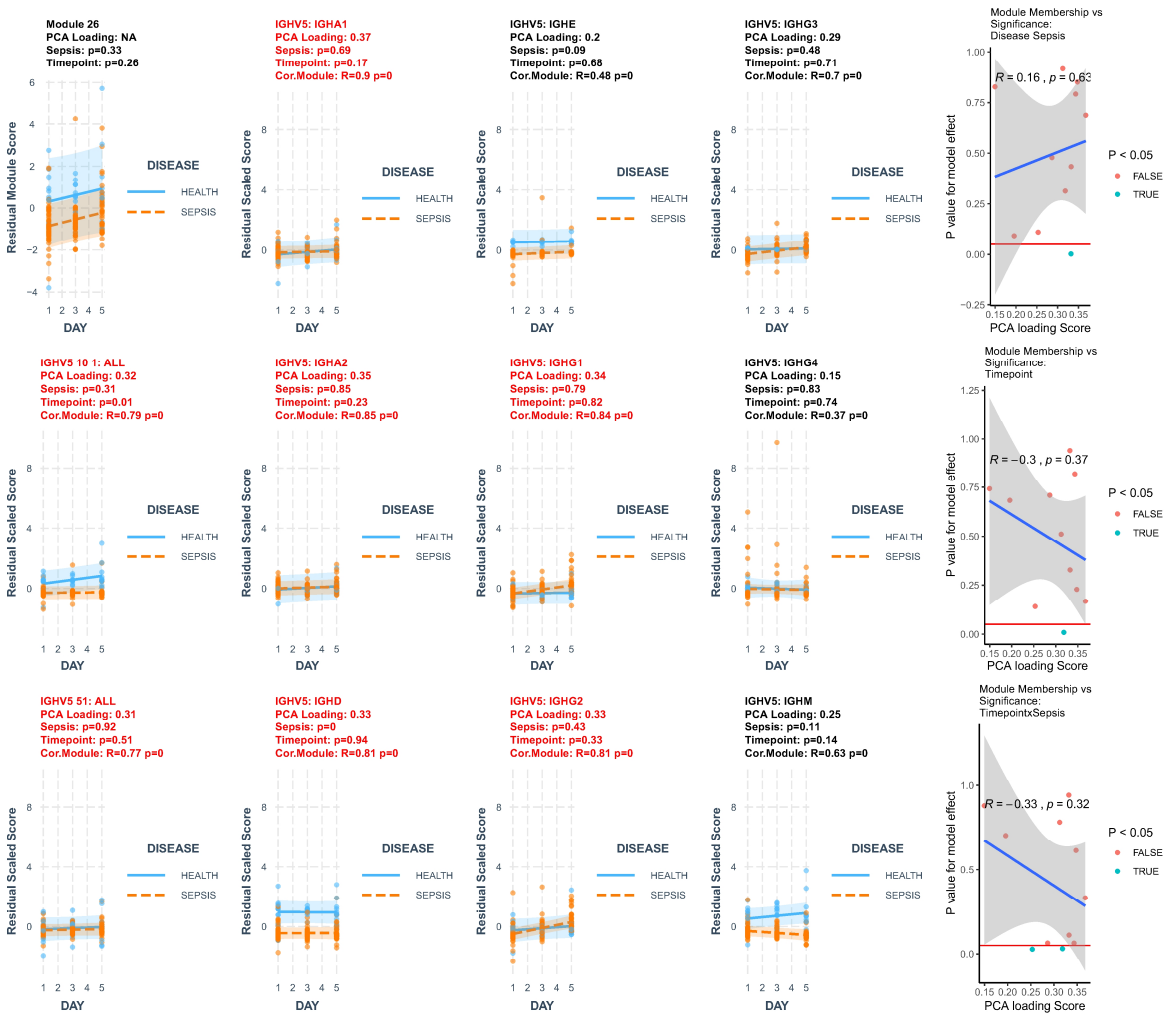


Figure B. 1. Partial residual plots showing the change in module 26 score over time in health and sepsis, controlling for age, sex and the Charlson Index.

Top left shows the partial residual plot for BCR module 26 using the D*TASC linear mixed effect model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 26 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). 95% confidence intervals are shown. Divergence in feature trajectories over time is seen for the IGHV5 usage for IGHM VDJ sequences.

	Linkage Method
Average linkage	Computes all pairwise dissimilarities between clusters 1 and 2 and considers the average value of these dissimilarities to be the distance between two clusters.
Weighted average linkage	Uses the same method as average linkage but weights distances based on the number of elements in the cluster.
Single linkage	Computes all pairwise dissimilarities between clusters 1 and 2 and considers the smallest value of these dissimilarities to be the distance between two clusters.
Complete linkage	Computes all pairwise dissimilarities between clusters 1 and 2 and considers the largest value of these dissimilarities to be the distance between two clusters.
Ward's linkage	Wards minimum variance method minimises the total within-cluster variance. The distance between two clusters is computed as the increase in 'error sum of squares' after merging the two clusters.

Table B. 6. Linkage methods.

Five different linkage methods were evaluated using the agglomeration coefficient (ac) score. The method with the highest ac score detects the most cluster structure and was used for downstream analysis.

	Cluster Partition Method
Dunn Index	Defines the ratio between the minimum inter-cluster distances to maximal intra-cluster index. The best cluster partition maximises the Dunn Index.
Silhouette Index	Calculates the similarity of an object to all objects within the same cluster and to all other objects across clusters. A high score indicates objects are well matched. Index is maximised at the optimal cluster partition.
Mcclain	A ratio of (the average within cluster distance divided by the number of within cluster distances) to the (between cluster distance divided by the number of cluster distances). The minimum value of the index indicates the optimal number of clusters.
C-index	Compares the within-cluster dissimilarity to the maximum within-cluster dissimilarity in the dataset. The optimal cluster partition minimises the C-index.
Frey	A ratio of scores from two successive levels of the hierarchy comparing the outside and within cluster distances at each of the two hierarchy levels. The best partition occurs when the ratio falls below 1 and the cluster level for the previous partition is taken as the optimal.

Table B. 7. Cluster partition methods using fixed height cutting, implemented in the NBclust R package.

Method definitions are sourced from Charrad et al. (2012)

$$Loading\ Score_i\ assuming\ equal\ contribution\ to\ PC1_i = \sqrt{\frac{1}{number\ of\ features\ in\ cluster_i}}$$

Equation B. 4. PC loading score assuming all features in a cluster contribute equally to PC1.

.

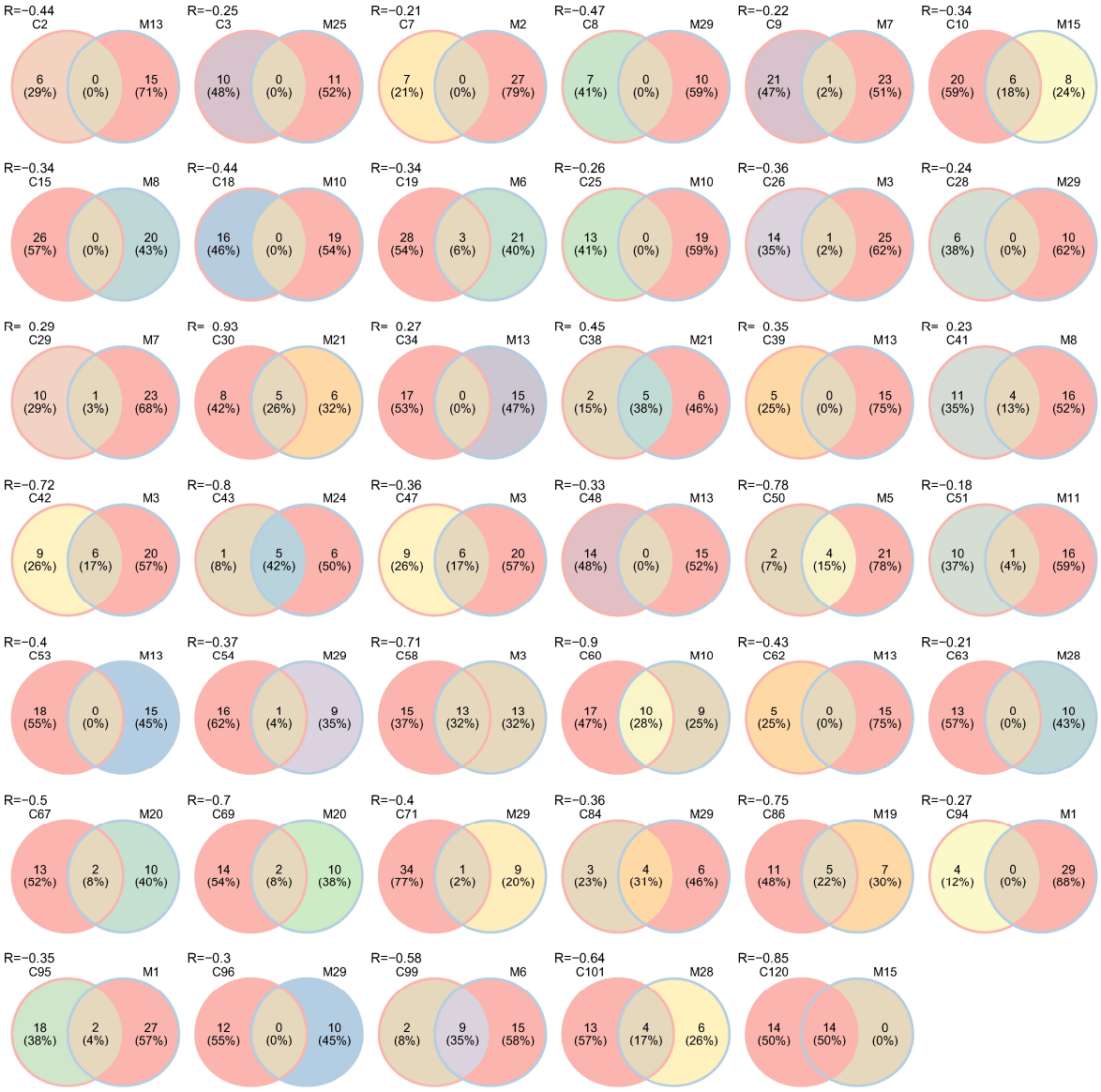


Figure B. 2. The number of overlapping BCR features between strongly negatively correlated SDA components (C) and rWGCNA modules (M).

The RMCORR coefficient is shown for each component and module pair. Modules and components with a large feature overlap are likely to capture similar biological signals.

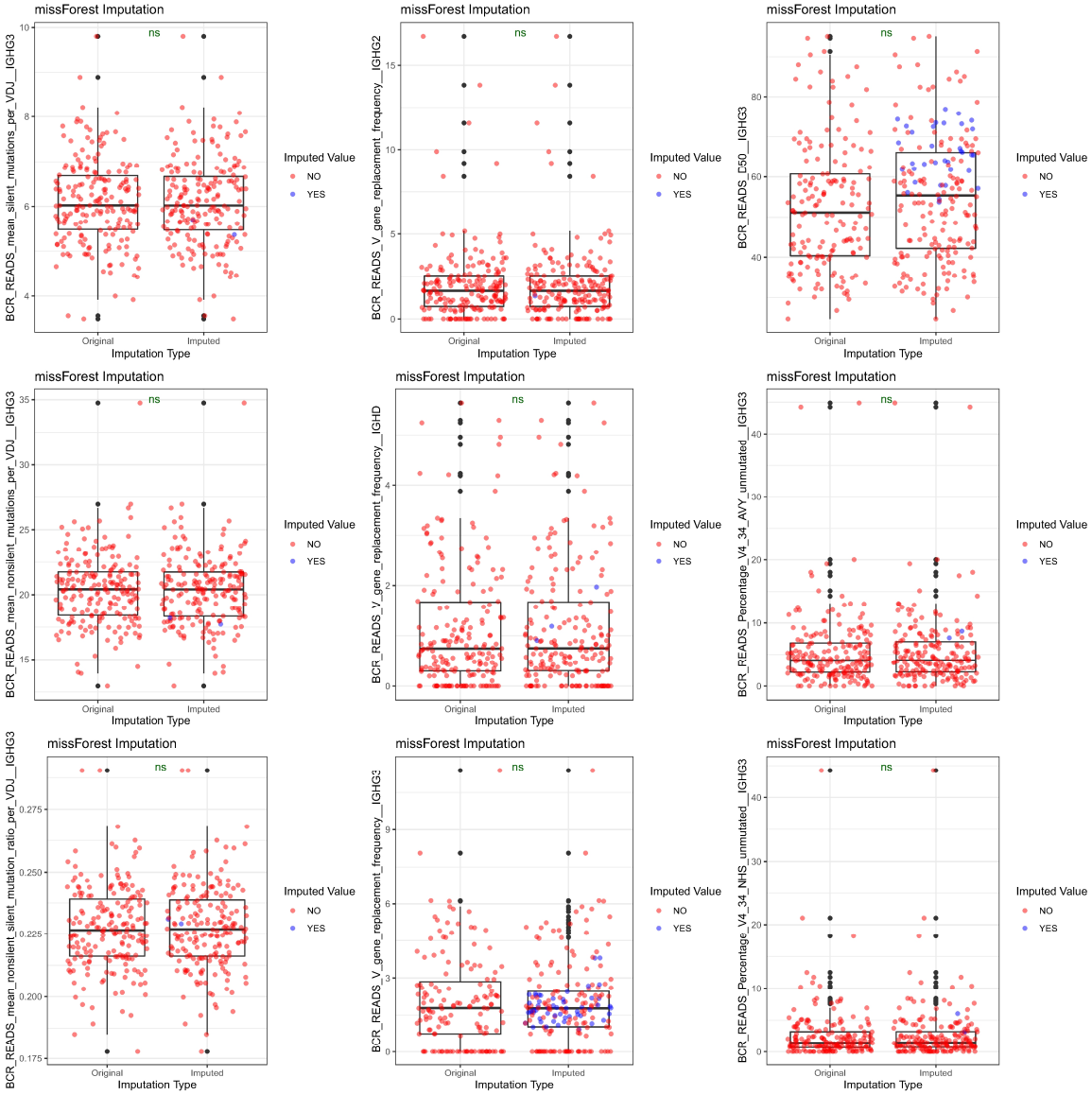


Figure B. 3. Distribution of a subset of BCR repertoire features pre and post-imputation with missForest.

Imputed values are shown in blue. The median, first and third quartile are shown and the whiskers represent 1.5x the interquartile range (IQR).

C Appendices: The BCR Repertoire in Sepsis

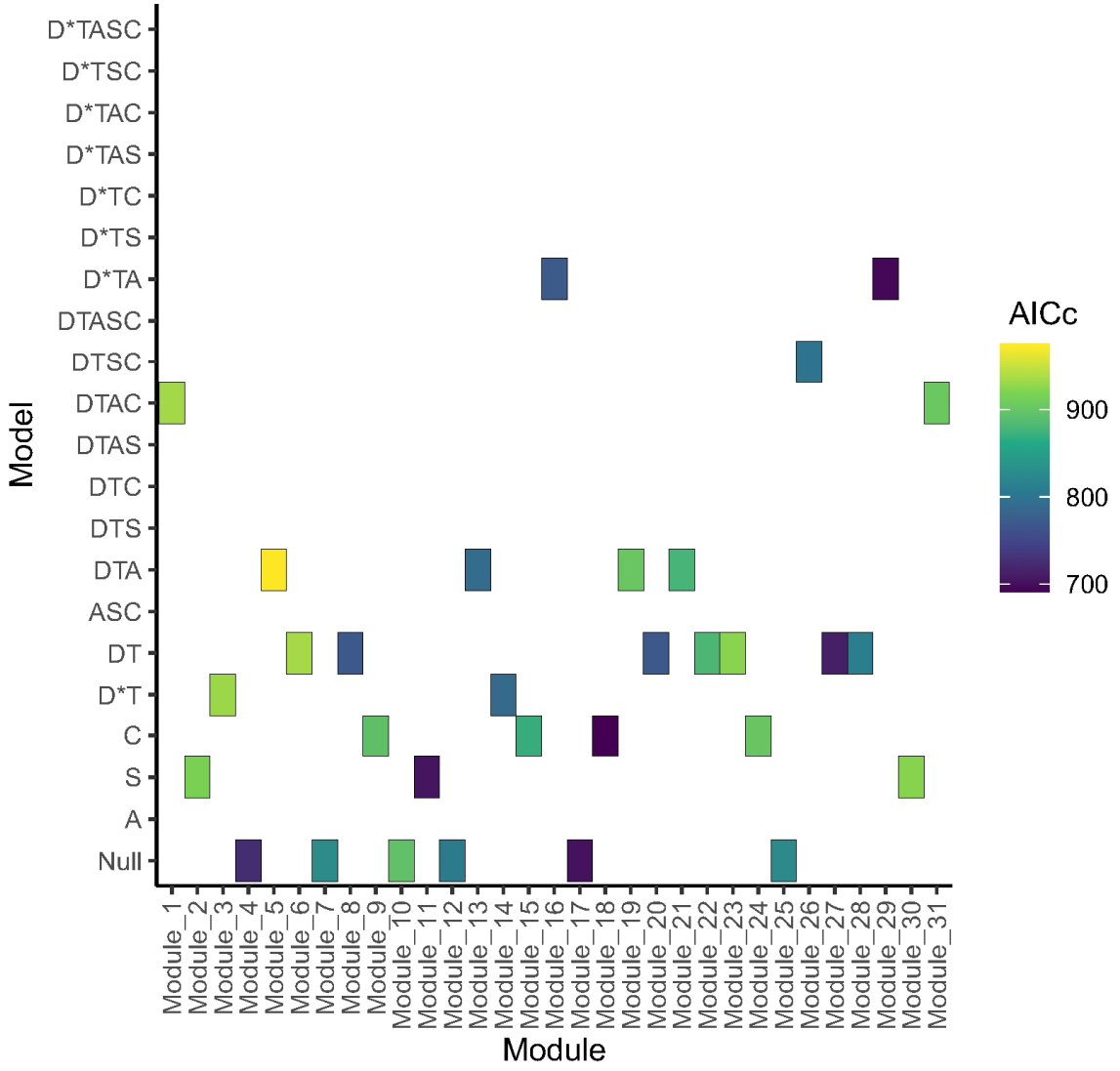


Figure C. 1. Optimal statistical model for BCR repertoire module analysis, determined by AICc score.

Module	Model	AICc	AICcWt
Module_1	DTAC	935.052	0.267
Module_2	S	917.840	0.309
Module_3	D*T	932.227	0.274
Module_4	Null	724.300	0.234
Module_5	DTA	976.429	0.193
Module_6	DT	936.599	0.213
Module_7	Null	828.532	0.274
Module_8	DT	769.468	0.223
Module_9	C	894.732	0.231
Module_10	Null	896.438	0.253
Module_11	S	703.323	0.250
Module_12	Null	808.391	0.243
Module_13	DTA	791.179	0.238
Module_14	D*T	789.700	0.296
Module_15	C	869.586	0.405
Module_16	D*TA	772.341	0.329
Module_17	Null	701.000	0.344
Module_18	C	690.483	0.242
Module_19	DTA	903.800	0.213
Module_20	DT	770.358	0.142
Module_21	DTA	878.968	0.227
Module_22	DT	882.650	0.171
Module_23	DT	924.648	0.201
Module_24	C	902.543	0.274
Module_25	Null	826.035	0.232
Module_26	DTSC	800.754	0.125
Module_27	DT	712.090	0.116
Module_28	DT	812.082	0.245
Module_29	D*TA	695.269	0.270
Module_30	S	924.501	0.213
Module_31	DTAC	905.403	0.143

Table C. 1. AIC based model selection.

AICc and weighted AICcWt scores are reported for the optimal statistical model for each BCR repertoire module. Model terms are recorded in **Table 23**. Scores are reported to 3 decimal places.

BCR Module	Model Term						ANOVA	
	DISEASE SEPSIS	TIME	TIME × SEPSIS	AGE	SEX MALE	CI	NULL MODEL	COV ONLY MODEL
Raw p values								
1	1.54E-02	4.66E-01	4.39E-01	1.01E-02	2.26E-01	1.38E-02	2.69E-04	2.68E-04
2	5.86E-01	6.11E-01	2.83E-01	8.14E-01	2.35E-02	7.73E-01	2.22E-01	3.61E-01
3	7.83E-03	7.69E-01	3.68E-05	8.48E-02	4.17E-01	2.88E-01	1.63E-19	2.59E-18
4	2.04E-01	7.56E-01	6.78E-01	1.12E-01	2.95E-01	4.22E-02	3.93E-01	5.98E-01
5	1.03E-01	9.32E-01	2.40E-01	2.83E-02	4.34E-01	3.72E-01	1.50E-02	3.24E-02
6	1.38E-01	7.08E-01	4.93E-01	2.46E-01	6.09E-01	1.99E-01	1.22E-01	3.76E-02
7	7.71E-01	9.33E-01	6.74E-01	4.89E-01	1.54E-01	9.18E-01	7.41E-01	7.30E-01
8	1.47E-01	4.04E-01	2.41E-01	9.89E-01	8.76E-01	5.42E-01	8.72E-05	2.40E-04
9	8.27E-01	5.58E-01	1.14E-01	4.98E-01	6.39E-01	2.60E-02	3.18E-03	1.41E-01
10	7.49E-01	7.07E-01	6.90E-01	2.05E-01	8.32E-01	2.73E-01	6.03E-01	4.34E-01
11	3.79E-01	1.37E-01	2.62E-01	5.64E-01	7.95E-02	9.63E-01	3.18E-01	4.52E-01
12	1.23E-01	6.57E-01	8.78E-01	3.31E-01	8.89E-01	3.18E-01	6.80E-01	4.00E-01
13	2.35E-02	6.54E-01	9.33E-01	3.54E-03	2.42E-01	2.09E-01	1.19E-03	6.40E-02
14	2.64E-01	5.99E-01	5.81E-03	6.70E-01	2.48E-01	9.10E-01	3.34E-04	2.78E-05
15	2.28E-01	4.69E-01	3.87E-01	7.59E-01	5.00E-01	5.45E-02	3.43E-01	6.37E-01
16	2.83E-02	3.22E-01	4.08E-02	3.18E-03	5.06E-01	6.83E-01	3.24E-04	2.96E-02
17	8.01E-01	5.29E-01	2.54E-01	5.80E-01	6.55E-01	5.44E-01	8.53E-01	5.44E-01
18	1.08E-01	4.27E-01	2.34E-01	9.40E-01	6.17E-01	4.53E-01	1.73E-01	3.87E-01
19	1.18E-02	4.99E-01	6.17E-01	6.20E-03	2.56E-01	1.76E-01	2.02E-03	5.22E-02
20	3.74E-01	4.46E-01	1.50E-01	2.73E-01	8.64E-01	9.29E-01	2.83E-05	5.17E-05
21	7.21E-02	3.03E-01	6.79E-01	6.76E-03	1.87E-01	2.20E-01	2.12E-04	3.14E-03
22	9.30E-02	9.20E-01	2.10E-01	1.83E-01	9.99E-01	6.48E-01	9.31E-04	2.50E-03
23	9.17E-02	5.96E-01	8.51E-01	1.59E-01	6.45E-01	2.56E-01	9.99E-02	6.09E-02
24	5.63E-01	6.99E-01	5.26E-01	3.14E-01	6.23E-01	7.22E-01	1.40E-01	6.04E-01
25	6.98E-01	3.12E-01	5.89E-01	2.32E-01	7.58E-01	8.42E-01	6.27E-01	5.41E-01
26	3.33E-01	2.63E-01	9.67E-01	4.20E-01	1.46E-01	8.86E-02	6.96E-02	7.03E-02
27	7.96E-01	8.59E-01	1.74E-01	1.10E-01	2.54E-01	2.02E-01	1.02E-01	7.83E-02
28	8.01E-02	1.25E-01	7.33E-01	5.74E-01	2.54E-01	7.58E-01	2.62E-04	2.48E-04
29	2.50E-04	9.16E-01	1.31E-03	1.05E-01	6.34E-01	5.38E-01	4.17E-16	1.76E-13
30	6.63E-01	6.70E-01	5.73E-01	5.00E-01	1.53E-01	7.82E-01	6.19E-01	9.50E-01
31	2.47E-02	9.03E-01	2.60E-01	1.17E-02	2.77E-01	9.32E-02	1.17E-02	2.78E-02
BH adjusted p values								
1	1.00E-01	7.65E-01	7.41E-01	7.81E-02	5.60E-01	9.54E-02	4.18E-03	4.18E-03
2	8.35E-01	8.35E-01	5.80E-01	9.05E-01	1.45E-01	8.75E-01	5.55E-01	6.73E-01
3	6.27E-02	8.75E-01	1.31E-03	3.45E-01	7.29E-01	5.85E-01	4.04E-17	3.21E-16
4	5.35E-01	8.72E-01	8.38E-01	3.87E-01	5.95E-01	2.09E-01	7.01E-01	8.35E-01
5	3.76E-01	9.56E-01	5.60E-01	1.56E-01	7.37E-01	6.86E-01	1.00E-01	1.71E-01
6	4.36E-01	8.49E-01	7.85E-01	5.60E-01	8.35E-01	5.35E-01	4.05E-01	1.94E-01
7	8.75E-01	9.56E-01	8.38E-01	7.85E-01	4.50E-01	9.56E-01	8.71E-01	8.65E-01
8	4.44E-01	7.10E-01	5.60E-01	9.93E-01	9.39E-01	8.13E-01	2.40E-03	4.18E-03
9	9.16E-01	8.27E-01	3.87E-01	7.85E-01	8.38E-01	1.54E-01	3.04E-02	4.36E-01
10	8.72E-01	8.49E-01	8.42E-01	5.35E-01	9.17E-01	5.70E-01	8.35E-01	7.37E-01
11	6.91E-01	4.36E-01	5.60E-01	8.27E-01	3.31E-01	9.75E-01	6.17E-01	7.50E-01
12	4.05E-01	8.38E-01	9.39E-01	6.31E-01	9.47E-01	6.17E-01	8.38E-01	7.08E-01
13	1.45E-01	8.38E-01	9.56E-01	3.25E-02	5.60E-01	5.36E-01	1.47E-02	2.94E-01
14	5.60E-01	8.35E-01	5.15E-02	8.38E-01	5.60E-01	9.56E-01	4.60E-03	1.17E-03
15	5.60E-01	7.65E-01	6.96E-01	8.72E-01	7.85E-01	2.60E-01	6.44E-01	8.38E-01
16	1.56E-01	6.19E-01	2.06E-01	3.04E-02	7.90E-01	8.38E-01	4.60E-03	1.59E-01
17	8.95E-01	8.13E-01	5.60E-01	8.35E-01	8.38E-01	8.13E-01	9.27E-01	8.13E-01
18	3.81E-01	7.35E-01	5.60E-01	9.59E-01	8.35E-01	7.50E-01	4.91E-01	6.96E-01
19	8.33E-02	7.85E-01	8.35E-01	5.31E-02	5.60E-01	4.92E-01	2.28E-02	2.54E-01
20	6.86E-01	7.48E-01	4.48E-01	5.70E-01	9.32E-01	9.56E-01	1.17E-03	1.60E-03
21	3.14E-01	6.05E-01	8.38E-01	5.59E-02	5.09E-01	5.55E-01	4.18E-03	3.04E-02
22	3.56E-01	9.56E-01	5.36E-01	5.04E-01	9.99E-01	8.38E-01	1.22E-02	2.69E-02
23	3.56E-01	8.35E-01	9.27E-01	4.59E-01	8.38E-01	5.60E-01	3.75E-01	2.85E-01
24	8.27E-01	8.45E-01	8.13E-01	6.17E-01	8.36E-01	8.60E-01	4.36E-01	8.35E-01
25	8.45E-01	6.17E-01	8.35E-01	5.60E-01	8.72E-01	9.24E-01	8.37E-01	8.13E-01
26	6.31E-01	5.60E-01	9.75E-01	7.29E-01	4.44E-01	3.54E-01	3.12E-01	3.12E-01
27	8.93E-01	9.30E-01	4.91E-01	3.83E-01	5.60E-01	5.35E-01	3.76E-01	3.31E-01
28	3.31E-01	4.07E-01	8.65E-01	8.32E-01	5.60E-01	8.72E-01	4.18E-03	4.18E-03
29	4.18E-03	9.56E-01	1.55E-02	3.78E-01	8.38E-01	8.13E-01	3.45E-14	1.09E-11
30	8.38E-01	8.38E-01	8.32E-01	7.85E-01	4.50E-01	8.82E-01	8.35E-01	9.65E-01
31	1.49E-01	9.56E-01	5.60E-01	8.33E-02	5.73E-01	3.56E-01	8.33E-02	1.56E-01

Table C. 2. Linear Mixed Effects Models: effect of disease status and time on BCR module score - raw and BH adjusted p values.

P values are reported for each main effect term in the health versus sepsis linear mixed effects models (D*TASC model). Additionally, p values are provided for the ANOVA test, which compares the variance explained by the D*TASC model to that explained by the null model or the covariate-only model. Significant p values ($p<0.05$) are denoted in red. P values are reported to 3 significant figures.

Module	Day	Model Term				ANOVA
		DISEASE SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL
Raw p values						
3	Day1	1.51E-03	5.38E-02	4.05E-01	2.56E-01	1.51E-03
	Day3	1.19E-03	5.53E-02	2.38E-01	2.06E-01	1.19E-03
	Day5	2.40E-05	8.42E-01	7.55E-01	9.25E-01	2.40E-05
14	Day1	4.89E-01	3.77E-01	4.93E-01	8.90E-01	4.89E-01
	Day3	2.65E-01	7.59E-01	3.78E-01	2.29E-01	2.65E-01
	Day5	9.47E-01	4.87E-01	5.34E-01	1.58E-01	9.47E-01
16	Day1	9.12E-03	1.20E-03	7.49E-01	3.03E-01	9.12E-03
	Day3	2.55E-01	4.81E-01	4.01E-01	6.63E-01	2.55E-01
	Day5	7.15E-01	2.04E-01	6.75E-02	7.57E-01	7.15E-01
29	Day1	5.24E-05	2.67E-02	4.84E-01	2.15E-01	5.24E-05
	Day3	4.53E-04	1.73E-01	5.44E-01	6.80E-01	4.53E-04
	Day5	9.87E-06	5.07E-01	8.55E-01	5.16E-01	9.87E-06
BH adjusted p values						
3	Day1	6.95E-03	1.84E-01	6.75E-01	5.12E-01	6.95E-03
	Day3	6.53E-03	1.84E-01	5.12E-01	5.12E-01	6.53E-03
	Day5	3.60E-04	9.16E-01	8.44E-01	9.47E-01	3.60E-04
14	Day1	7.03E-01	6.66E-01	7.03E-01	9.36E-01	7.03E-01
	Day3	5.12E-01	8.44E-01	6.66E-01	5.12E-01	5.12E-01
	Day5	9.47E-01	7.03E-01	7.10E-01	4.75E-01	9.47E-01
16	Day1	3.65E-02	6.53E-03	8.44E-01	5.68E-01	3.65E-02
	Day3	5.12E-01	7.03E-01	6.75E-01	8.44E-01	5.12E-01
	Day5	8.44E-01	5.12E-01	2.13E-01	8.44E-01	8.44E-01
29	Day1	5.24E-04	1.00E-01	7.03E-01	5.12E-01	5.24E-04
	Day3	3.40E-03	4.95E-01	7.10E-01	8.44E-01	3.40E-03
	Day5	2.96E-04	7.03E-01	9.16E-01	7.03E-01	2.96E-04

Table C. 3. Simple Effects Tests: effect of disease status (health or sepsis) at each time point on BCR module score (DASC model) – raw and BH adjusted p values.

P values are reported for each model term and also for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are denoted in red. All p values are reported to 3 significant figures.

Module	Disease Group	Simple Effects Tests: Effect of time in health or sepsis				ANOVA	
		TIME	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values							
3	Health	5.12E-01	3.93E-01	6.07E-01	0.00E+00	5.07E-01	7.32E-01
	Sepsis	6.87E-14	4.33E-02	3.45E-01	1.95E-01	3.74E-13	5.12E-12
14	Health	3.84E-01	9.50E-01	2.50E-01	0.00E+00	3.79E-01	5.29E-01
	Sepsis	6.92E-06	6.62E-01	3.94E-01	8.85E-01	8.30E-06	3.96E-04
16	Health	3.61E-01	5.43E-01	5.12E-01	0.00E+00	3.58E-01	6.02E-01
	Sepsis	7.91E-03	4.83E-03	5.94E-01	5.68E-01	7.47E-03	1.69E-03
29	Health	8.95E-01	4.55E-01	5.82E-01	0.00E+00	8.93E-01	8.65E-01
	Sepsis	4.12E-09	3.09E-02	4.49E-01	3.19E-01	7.12E-09	6.30E-08
BH adjusted p values							
3	Health	7.22E-01	6.94E-01	7.22E-01	0.00E+00	7.22E-01	8.26E-01
	Sepsis	3.02E-12	1.27E-01	6.94E-01	5.36E-01	8.22E-12	7.50E-11
14	Health	6.94E-01	9.50E-01	6.48E-01	0.00E+00	6.94E-01	7.22E-01
	Sepsis	4.35E-05	7.67E-01	6.94E-01	9.15E-01	4.57E-05	1.93E-03
16	Health	6.94E-01	7.22E-01	7.22E-01	0.00E+00	6.94E-01	7.22E-01
	Sepsis	2.68E-02	1.93E-02	7.22E-01	7.22E-01	2.68E-02	7.44E-03
29	Health	9.15E-01	7.22E-01	7.22E-01	0.00E+00	9.15E-01	9.15E-01
	Sepsis	4.53E-08	9.71E-02	7.22E-01	6.94E-01	6.27E-08	4.62E-07

Table C. 4. Simple Effects Tests: effect of time on health or sepsis BCR module score (TASC model) – raw and BH adjusted p values.

P values are reported for each model term and also for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only and null models. Significant p values ($p<0.05$) are denoted in red. All p values are reported to 3 significant figures.

Module	Linear Mixed Effects Model								ANOVA	
	EARLY DEATH	LATE DEATH	TIME	TIME x EARLY DEATH	TIME x LATE DEATH	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values										
1	3.82E-01	4.80E-01	1.25E-02	3.60E-01	1.28E-01	7.04E-03	1.65E-01	1.14E-02	2.75E-02	3.97E-03
2	8.93E-02	7.93E-01	1.64E-01	5.40E-01	4.27E-01	7.27E-01	1.37E-01	9.60E-01	1.22E-01	1.95E-01
3	2.00E-04	2.18E-01	2.62E-14	4.39E-04	4.55E-01	5.12E-02	2.74E-01	1.56E-01	9.58E-14	2.85E-13
4	3.97E-02	4.80E-01	5.03E-01	8.30E-02	8.54E-01	1.33E-01	4.26E-01	4.28E-02	4.33E-01	2.49E-01
5	1.53E-01	9.57E-01	2.57E-01	8.25E-02	8.49E-01	3.14E-02	3.51E-01	3.27E-01	1.12E-01	7.74E-02
6	7.44E-01	4.91E-01	4.12E-02	3.89E-01	6.77E-01	4.26E-01	4.21E-01	2.81E-01	1.81E-01	3.17E-01
7	3.76E-01	8.26E-01	8.59E-02	2.90E-01	6.09E-01	8.40E-01	2.95E-01	9.66E-01	3.75E-01	5.79E-01
8	3.16E-02	3.34E-01	3.57E-04	6.58E-01	7.19E-01	8.37E-01	9.60E-01	6.50E-01	2.44E-04	2.09E-03
9	1.62E-05	2.85E-02	2.68E-05	1.54E-04	3.95E-02	3.22E-01	9.33E-01	2.88E-02	3.36E-05	5.09E-05
10	3.54E-01	4.98E-01	2.14E-01	1.39E-01	5.18E-01	1.58E-01	5.06E-01	1.41E-01	1.02E-01	1.45E-01
11	1.94E-01	9.21E-01	1.52E-01	3.22E-01	1.95E-01	4.91E-01	2.02E-01	7.93E-01	4.76E-01	5.87E-01
12	1.37E-01	5.92E-01	1.90E-01	1.12E-01	5.34E-01	2.68E-01	8.56E-01	2.64E-01	6.10E-01	7.20E-01
13	8.19E-01	8.09E-01	6.83E-01	7.80E-01	7.09E-01	4.22E-03	1.42E-01	1.67E-01	9.85E-01	2.30E-01
14	2.04E-02	4.60E-01	1.35E-06	1.21E-02	3.19E-01	6.76E-01	4.95E-01	8.85E-01	5.89E-05	5.62E-04
15	8.91E-02	6.59E-02	1.55E-01	3.65E-02	4.47E-01	5.99E-01	5.05E-01	4.61E-02	1.81E-01	1.31E-01
16	1.15E-03	9.77E-01	9.08E-01	1.68E-04	3.52E-01	3.40E-03	8.32E-01	5.53E-01	2.60E-04	4.54E-05
17	9.51E-02	2.08E-01	5.31E-02	1.64E-06	2.16E-01	1.61E-01	2.53E-02	4.94E-01	2.13E-06	8.88E-06
18	3.27E-01	6.98E-01	8.26E-01	8.79E-03	4.46E-01	8.80E-01	8.52E-01	3.76E-01	7.19E-02	2.13E-01
19	8.71E-01	9.45E-01	8.29E-01	5.64E-01	8.66E-01	5.34E-03	1.20E-01	1.48E-01	9.88E-01	2.70E-01
20	1.16E-03	2.47E-01	3.52E-06	4.05E-03	5.05E-01	1.75E-01	4.37E-01	8.08E-01	4.38E-05	1.48E-04
21	8.85E-01	6.50E-01	1.25E-01	2.61E-01	6.20E-01	5.55E-03	1.21E-01	1.15E-01	7.37E-02	8.95E-03
22	3.42E-02	8.68E-01	2.41E-02	7.88E-02	2.07E-01	2.16E-01	9.28E-01	5.58E-01	1.17E-02	3.41E-02
23	7.74E-02	9.47E-01	6.22E-01	1.05E-01	6.36E-01	1.87E-01	8.22E-01	2.93E-01	2.71E-01	4.14E-01
24	7.86E-01	8.47E-01	6.44E-01	9.56E-01	8.17E-01	3.88E-01	3.96E-01	7.13E-01	9.87E-01	9.31E-01
25	5.30E-02	1.38E-01	7.08E-03	1.10E-03	1.46E-01	4.62E-01	9.76E-01	7.17E-01	1.88E-02	5.72E-02
26	2.78E-01	6.88E-01	2.55E-01	3.74E-03	4.41E-01	3.82E-01	5.08E-01	1.91E-01	3.60E-04	8.94E-04
27	3.74E-01	9.85E-01	8.26E-02	8.95E-01	5.31E-01	3.13E-01	1.44E-01	2.74E-01	1.95E-01	2.04E-01
28	5.15E-03	8.94E-02	2.08E-06	1.70E-02	1.82E-02	4.02E-01	4.46E-01	5.44E-01	2.86E-04	1.57E-03
29	1.37E-03	2.03E-01	9.67E-10	1.72E-03	5.84E-01	3.83E-02	3.88E-01	2.89E-01	5.41E-09	1.17E-08
30	6.98E-01	3.46E-01	9.31E-01	5.21E-01	2.86E-01	4.48E-01	6.98E-02	7.14E-01	5.32E-01	4.53E-01
31	2.18E-01	2.86E-01	3.00E-02	5.07E-01	5.09E-01	1.04E-02	2.10E-01	9.51E-02	2.92E-01	1.18E-01
BH Adjusted values										
1	6.23E-01	6.96E-01	7.04E-02	6.07E-01	3.70E-01	4.57E-02	4.02E-01	6.78E-02	1.38E-01	2.99E-02
2	2.95E-01	9.17E-01	4.02E-01	7.22E-01	6.61E-01	8.56E-01	3.81E-01	9.82E-01	3.59E-01	4.39E-01
3	2.70E-03	4.51E-01	8.11E-12	4.70E-03	6.74E-01	2.06E-01	5.25E-01	3.93E-01	1.48E-11	2.94E-11
4	1.68E-01	6.96E-01	6.98E-01	2.86E-01	9.35E-01	3.77E-01	6.61E-01	1.77E-01	6.68E-01	5.05E-01
5	3.92E-01	9.82E-01	5.15E-01	2.86E-01	9.35E-01	1.46E-01	5.99E-01	5.70E-01	3.49E-01	2.79E-01
6	8.73E-01	6.98E-01	1.73E-01	6.24E-01	8.33E-01	6.61E-01	6.61E-01	5.32E-01	4.29E-01	5.67E-01
7	6.20E-01	9.31E-01	2.92E-01	5.34E-01	7.78E-01	9.33E-01	5.35E-01	9.85E-01	6.20E-01	7.57E-01
8	1.46E-01	5.79E-01	3.99E-03	8.16E-01	8.52E-01	9.33E-01	9.82E-01	8.09E-01	3.16E-03	1.71E-02
9	3.87E-04	1.40E-01	5.94E-04	2.27E-03	1.68E-01	5.68E-01	9.73E-01	1.40E-01	6.94E-04	8.77E-04
10	5.99E-01	6.98E-01	4.51E-01	3.82E-01	7.07E-01	3.94E-01	6.98E-01	3.83E-01	3.25E-01	3.85E-01
11	4.39E-01	9.73E-01	3.92E-01	5.68E-01	4.39E-01	6.98E-01	4.49E-01	9.17E-01	6.96E-01	7.61E-01
12	3.81E-01	7.65E-01	4.39E-01	3.49E-01	7.17E-01	5.25E-01	9.35E-01	5.20E-01	7.78E-01	8.52E-01
13	9.31E-01	9.29E-01	8.36E-01	9.12E-01	8.52E-01	3.04E-02	3.83E-01	4.04E-01	9.88E-01	4.71E-01
14	1.07E-01	6.79E-01	5.99E-05	6.95E-02	5.68E-01	8.33E-01	6.98E-01	9.46E-01	9.61E-04	5.81E-03
15	2.95E-01	2.52E-01	3.93E-01	1.62E-01	6.71E-01	7.71E-01	6.98E-01	1.88E-01	4.29E-01	3.75E-01
16	1.06E-02	9.88E-01	9.64E-01	2.36E-03	5.99E-01	2.70E-02	9.31E-01	7.33E-01	3.23E-03	8.28E-04
17	3.07E-01	4.51E-01	2.08E-01	6.34E-05	4.51E-01	4.00E-01	1.28E-01	6.98E-01	6.60E-05	2.30E-04
18	5.70E-01	8.46E-01	9.31E-01	5.55E-02	6.71E-01	9.46E-01	9.35E-01	6.20E-01	2.69E-01	4.51E-01
19	9.41E-01	9.82E-01	9.31E-01	7.41E-01	9.40E-01	3.68E-02	3.59E-01	3.86E-01	9.88E-01	5.25E-01
20	1.06E-02	5.04E-01	9.93E-05	2.99E-02	6.98E-01	4.21E-01	6.71E-01	9.29E-01	8.28E-04	2.27E-03
21	9.46E-01	8.09E-01	3.65E-01	5.18E-01	7.87E-01	3.74E-02	3.59E-01	3.53E-01	2.72E-01	5.55E-02
22	1.54E-01	9.40E-01	1.25E-01	2.81E-01	4.51E-01	4.51E-01	9.73E-01	7.36E-01	6.87E-02	1.54E-01
23	2.79E-01	9.82E-01	7.87E-01	3.33E-01	8.01E-01	4.39E-01	9.31E-01	5.34E-01	5.25E-01	6.55E-01
24	9.16E-01	9.35E-01	8.08E-01	9.82E-01	9.31E-01	6.24E-01	6.33E-01	8.52E-01	9.88E-01	9.73E-01
25	2.08E-01	3.81E-01	4.57E-02	1.06E-02	3.85E-01	6.79E-01	9.88E-01	8.52E-01	1.00E-01	2.22E-01
26	5.29E-01	8.39E-01	5.12E-01	2.90E-02	6.71E-01	6.23E-01	6.98E-01	4.39E-01	3.99E-03	8.94E-03
27	6.20E-01	9.88E-01	2.86E-01	9.53E-01	7.17E-01	5.64E-01	3.85E-01	5.25E-01	4.39E-01	4.49E-01
28	3.63E-02	2.95E-01	6.60E-05	9.42E-02	9.91E-02	6.39E-01	6.71E-01	7.24E-01	3.42E-03	1.35E-02
29	1.21E-02	4.49E-01	7.49E-08	1.44E-02	7.61E-01	1.67E-01	6.24E-01	5.34E-01	3.35E-07	6.05E-07
30	8.46E-01	5.95E-01	9.73E-01	7.08E-01	5.34E-01	6.71E-01	2.64E-01	8.52E-01	7.17E-01	6.74E-01
31	4.51E-01	5.34E-01	1.43E-01	6.98E-01	6.98E-01	6.35E-02	4.51E-01	3.07E-01	5.34E-01	3.57E-01

Table C. 5. Linear Mixed Effects Models: effect of sepsis mortality group and time on BCR module score - raw and BH adjusted p values. P values are reported for each main effect term in the health versus sepsis linear mixed effects models and for the ANOVA test, which compares the variance explained by the complete model to that explained by the null model or the covariate-only model. Significant p values ($p<0.05$) are denoted in red. All p values are reported to 3 significant figures.

Module	Day	Simple Effects Tests: Effect of mortality at each time point					ANOVA
		MORTALITY LATE DEATH	MORTALITY EARLY DEATH	AGE	SEX MALE	CI	COV ONLY MODEL
3	Day1	4.21E-01	1.29E-02	4.21E-01	4.68E-02	1.21E-01	4.38E-02
	Day3	1.15E-01	1.71E-01	2.27E-01	3.66E-02	9.28E-02	1.86E-01
	Day5	6.74E-01	4.90E-01	4.93E-01	6.25E-01	8.00E-01	6.64E-01
9	Day1	9.93E-02	1.15E-03	8.02E-01	3.65E-01	2.93E-02	4.44E-03
	Day3	8.00E-02	8.11E-02	6.81E-01	9.80E-01	3.58E-01	9.74E-02
	Day5	7.95E-01	8.63E-01	6.39E-01	3.34E-01	4.87E-01	9.40E-01
14	Day1	4.53E-01	8.41E-02	6.42E-01	5.59E-01	8.19E-01	2.21E-01
	Day3	8.71E-01	8.24E-01	3.39E-01	3.82E-01	1.39E-01	9.47E-01
	Day5	7.07E-01	1.98E-01	9.50E-01	4.29E-01	1.47E-01	4.28E-01
15	Day1	1.38E-01	3.58E-01	3.44E-01	4.43E-01	7.88E-02	3.04E-01
	Day3	3.36E-02	1.91E-01	9.75E-01	8.97E-01	1.69E-01	8.19E-02
	Day5	2.94E-01	6.01E-01	5.33E-01	3.72E-01	1.78E-01	4.30E-01
16	Day1	8.62E-01	1.08E-02	8.96E-01	9.22E-04	1.15E-01	1.95E-02
	Day3	3.70E-01	1.74E-01	5.29E-01	4.74E-01	4.94E-01	3.46E-01
	Day5	2.37E-01	1.38E-01	3.57E-01	1.40E-01	8.77E-01	2.25E-01
17	Day1	1.94E-01	9.88E-01	1.03E-02	4.22E-01	8.48E-01	3.78E-01
	Day3	9.15E-01	5.30E-04	4.35E-02	8.99E-01	5.34E-01	1.50E-03
	Day5	9.32E-01	4.00E-03	6.19E-01	2.11E-02	4.95E-02	1.20E-02
18	Day1	8.70E-01	6.31E-01	3.85E-01	4.86E-01	6.56E-01	8.32E-01
	Day3	7.37E-01	1.55E-02	1.82E-01	6.91E-01	5.82E-01	2.92E-02
	Day5	3.23E-01	5.62E-02	5.87E-01	2.82E-01	3.31E-01	1.34E-01
20	Day1	3.70E-01	2.48E-02	4.56E-01	1.02E-01	4.86E-01	7.87E-02
	Day3	1.44E-01	7.32E-02	3.78E-01	1.66E-01	5.64E-01	1.27E-01
	Day5	8.68E-01	3.00E-01	6.79E-01	5.06E-01	3.40E-01	5.38E-01
25	Day1	3.01E-01	3.75E-01	7.99E-01	3.47E-01	7.14E-01	5.03E-01
	Day3	3.86E-01	9.90E-01	8.05E-01	2.15E-01	8.73E-01	6.56E-01
	Day5	6.55E-01	9.03E-02	6.62E-01	7.54E-01	5.46E-01	1.70E-01
26	Day1	8.17E-01	4.98E-02	2.94E-01	4.17E-01	3.93E-01	8.69E-02
	Day3	8.12E-01	4.67E-02	5.00E-01	5.43E-01	2.54E-01	9.38E-02
	Day5	6.88E-01	4.48E-03	5.60E-01	7.21E-01	6.27E-01	1.62E-02
28	Day1	3.37E-01	2.71E-02	7.68E-01	4.72E-01	3.96E-01	8.44E-02
	Day3	2.94E-01	5.11E-01	1.39E-01	4.68E-01	4.02E-01	5.38E-01
	Day5	3.00E-01	8.05E-01	6.29E-02	7.98E-01	9.35E-01	5.79E-01
29	Day1	3.78E-01	2.36E-02	4.76E-01	1.86E-02	9.28E-02	7.53E-02
	Day3	1.47E-01	3.62E-01	3.39E-01	8.82E-02	3.99E-01	3.06E-01
	Day5	3.77E-01	9.47E-01	5.91E-01	8.70E-01	7.51E-01	6.68E-01

Table C. 6. Simple Effects Tests: effect of sepsis mortality group at each time point on BCR module score – raw p values.

P values are reported for each main effect term and also reported for the results of the ANOVA test, comparing the variance explained by the MASC model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

Module	Day	Simple Effects Tests: Effect of Mortality at each time point					ANOVA
		MORTALITY LATE DEATH	MORTALITY EARLY DEATH	AGE	SEX MALE	CI	COV ONLY MODEL
3	Day1	7.43E-01	2.53E-01	7.43E-01	3.71E-01	5.01E-01	3.71E-01
	Day3	4.86E-01	5.59E-01	6.37E-01	3.44E-01	4.41E-01	5.74E-01
	Day5	8.55E-01	7.73E-01	7.73E-01	8.47E-01	9.25E-01	8.54E-01
9	Day1	4.47E-01	8.10E-02	9.25E-01	7.31E-01	3.01E-01	1.38E-01
	Day3	4.41E-01	4.41E-01	8.55E-01	9.89E-01	7.31E-01	4.47E-01
	Day5	9.25E-01	9.38E-01	8.51E-01	7.31E-01	7.73E-01	9.68E-01
14	Day1	7.69E-01	4.41E-01	8.51E-01	8.01E-01	9.26E-01	6.35E-01
	Day3	9.38E-01	9.27E-01	7.31E-01	7.31E-01	5.12E-01	9.68E-01
	Day5	8.73E-01	5.85E-01	9.68E-01	7.43E-01	5.12E-01	7.43E-01
15	Day1	5.12E-01	7.31E-01	7.31E-01	7.59E-01	4.41E-01	7.31E-01
	Day3	3.30E-01	5.81E-01	9.89E-01	9.47E-01	5.59E-01	4.41E-01
	Day5	7.31E-01	8.26E-01	7.90E-01	7.31E-01	5.66E-01	7.43E-01
16	Day1	9.38E-01	2.53E-01	9.47E-01	8.10E-02	4.86E-01	2.80E-01
	Day3	7.31E-01	5.60E-01	7.90E-01	7.73E-01	7.73E-01	7.31E-01
	Day5	6.56E-01	5.12E-01	7.31E-01	5.12E-01	9.38E-01	6.37E-01
17	Day1	5.82E-01	9.90E-01	2.53E-01	7.43E-01	9.38E-01	7.31E-01
	Day3	9.59E-01	8.10E-02	3.71E-01	9.47E-01	7.90E-01	8.10E-02
	Day5	9.68E-01	1.38E-01	8.46E-01	2.85E-01	3.71E-01	2.53E-01
18	Day1	9.38E-01	8.47E-01	7.31E-01	7.73E-01	8.53E-01	9.31E-01
	Day3	8.94E-01	2.69E-01	5.71E-01	8.58E-01	8.16E-01	3.01E-01
	Day5	7.31E-01	4.05E-01	8.18E-01	7.31E-01	7.31E-01	5.12E-01
20	Day1	7.31E-01	2.98E-01	7.69E-01	4.50E-01	7.73E-01	4.41E-01
	Day3	5.12E-01	4.41E-01	7.31E-01	5.59E-01	8.01E-01	5.12E-01
	Day5	9.38E-01	7.31E-01	8.55E-01	7.75E-01	7.31E-01	7.90E-01
25	Day1	7.31E-01	7.31E-01	9.25E-01	7.31E-01	8.76E-01	7.75E-01
	Day3	7.31E-01	9.90E-01	9.25E-01	6.28E-01	9.38E-01	8.53E-01
	Day5	8.53E-01	4.41E-01	8.54E-01	9.05E-01	7.92E-01	5.59E-01
26	Day1	9.26E-01	3.71E-01	7.31E-01	7.43E-01	7.36E-01	4.41E-01
	Day3	9.26E-01	3.71E-01	7.75E-01	7.92E-01	6.94E-01	4.41E-01
	Day5	8.58E-01	1.38E-01	8.01E-01	8.80E-01	8.47E-01	2.69E-01
28	Day1	7.31E-01	3.01E-01	9.17E-01	7.73E-01	7.36E-01	4.41E-01
	Day3	7.31E-01	7.78E-01	5.12E-01	7.73E-01	7.36E-01	7.90E-01
	Day5	7.31E-01	9.25E-01	4.39E-01	9.25E-01	9.68E-01	8.16E-01
29	Day1	7.31E-01	2.98E-01	7.73E-01	2.80E-01	4.41E-01	4.41E-01
	Day3	5.12E-01	7.31E-01	7.31E-01	4.41E-01	7.36E-01	7.31E-01
	Day5	7.31E-01	9.68E-01	8.18E-01	9.38E-01	9.05E-01	8.54E-01

Table C. 7. Simple Effects Tests: effect of sepsis mortality group at each time point on BCR module score BH adjusted p values.

BH adjusted p values are reported for each main effect term (MASC model). BH Adjusted p values are also reported for the results of the ANOVA test, comparing the variance explained by the MASC model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

Module	Mortality Group	Simple Effects Tests: Effect of Time				ANOVA	
		TIMEPOINT	AGE	SEX MALE	COMORBIDITIES CI	COV ONLY MODEL	NULL MODEL
3	EARLY.DEATH	7.65E-01	1.99E-03	1.26E-02	2.25E-01	7.67E-01	1.36E-02
	LATE.DEATH	1.13E-04	6.92E-01	6.70E-01	2.33E-01	9.37E-05	1.16E-03
	SURVIVORS	1.29E-11	6.88E-01	8.97E-01	1.00E+00	1.85E-11	2.59E-09
9	EARLY.DEATH	1.04E-01	4.18E-01	8.43E-01	1.04E-03	9.93E-02	9.31E-03
	LATE.DEATH	6.78E-01	5.36E-01	3.69E-01	7.10E-01	6.74E-01	8.18E-01
	SURVIVORS	2.41E-05	7.50E-01	8.24E-01	3.47E-01	2.37E-05	7.74E-04
14	EARLY.DEATH	7.90E-01	1.95E-01	2.11E-01	3.75E-02	7.87E-01	1.87E-01
	LATE.DEATH	2.86E-02	9.53E-01	1.13E-01	5.97E-01	2.73E-02	1.30E-01
	SURVIVORS	2.44E-05	3.36E-01	8.16E-01	1.62E-01	2.09E-05	5.16E-04
15	EARLY.DEATH	1.18E-01	3.37E-01	4.03E-02	3.88E-02	1.05E-01	3.21E-02
	LATE.DEATH	1.00E+00	9.98E-01	8.77E-01	8.91E-01	1.00E+00	1.00E+00
	SURVIVORS	1.27E-01	8.71E-02	5.28E-01	2.64E-02	1.26E-01	1.14E-01
16	EARLY.DEATH	6.09E-04	3.39E-02	4.73E-01	1.86E-01	6.57E-04	8.40E-04
	LATE.DEATH	1.35E-01	3.26E-01	4.77E-02	4.82E-01	1.31E-01	1.19E-01
	SURVIVORS	9.35E-01	1.80E-01	2.53E-01	3.56E-01	9.35E-01	5.85E-02
17	EARLY.DEATH	1.88E-03	3.10E-01	3.30E-01	3.20E-01	1.24E-03	1.79E-02
	LATE.DEATH	5.65E-03	8.91E-01	6.30E-01	3.81E-03	5.27E-03	9.94E-04
	SURVIVORS	1.64E-02	3.17E-01	5.34E-02	4.58E-01	1.60E-02	3.29E-02
18	EARLY.DEATH	1.75E-02	8.56E-01	1.32E-01	5.89E-01	1.53E-02	3.34E-02
	LATE.DEATH	4.15E-01	5.41E-01	7.67E-01	5.68E-01	4.09E-01	6.99E-01
	SURVIVORS	8.21E-01	5.03E-01	2.31E-01	1.76E-01	8.21E-01	4.48E-01
20	EARLY.DEATH	4.20E-01	9.64E-03	4.30E-01	3.46E-01	4.16E-01	1.10E-01
	LATE.DEATH	1.21E-02	9.06E-01	6.79E-01	5.13E-01	1.11E-02	1.14E-01
	SURVIVORS	3.27E-05	5.19E-01	1.08E-01	7.39E-02	2.96E-05	7.91E-05
25	EARLY.DEATH	7.14E-02	1.55E-01	2.06E-01	2.12E-01	7.00E-02	1.07E-01
	LATE.DEATH	9.94E-01	9.12E-01	2.36E-01	9.08E-02	9.94E-01	1.30E-01
	SURVIVORS	1.90E-02	5.89E-02	2.05E-01	6.55E-01	1.83E-02	1.69E-02
26	EARLY.DEATH	4.03E-02	8.57E-01	3.44E-01	2.54E-01	4.78E-02	1.14E-01
	LATE.DEATH	5.28E-02	3.35E-01	9.73E-02	3.60E-03	4.99E-02	2.18E-03
	SURVIVORS	1.72E-01	2.24E-01	4.44E-01	8.28E-01	1.71E-01	1.98E-01
28	EARLY.DEATH	6.38E-01	2.76E-02	7.28E-01	8.64E-01	6.34E-01	2.14E-01
	LATE.DEATH	7.94E-01	1.58E-01	1.25E-01	4.30E-01	7.92E-01	4.20E-01
	SURVIVORS	3.09E-06	8.08E-01	8.02E-01	3.16E-01	3.35E-06	1.05E-04
29	EARLY.DEATH	7.46E-01	7.40E-04	6.59E-02	3.34E-01	7.43E-01	1.44E-02
	LATE.DEATH	1.78E-03	5.08E-01	5.10E-01	5.32E-01	1.57E-03	1.92E-02
	SURVIVORS	2.29E-08	6.17E-01	8.16E-01	9.64E-01	2.74E-08	2.11E-06

Table C. 8. Simple Effects Tests: effect of time on BCR module for sepsis mortality groups - raw p values.

P values are reported for each main effect term and for the results of the ANOVA test, comparing the variance explained by the TASC model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to three decimal places.

Module	Mortality Group	Simple Effects Tests: Effect of Time				ANOVA	
		TIMEPOINT	AGE	SEX MALE	COMORBIDITIES CI	COV ONLY MODEL	NULL MODEL
3	EARLY.DEATH	9.11E-01	1.34E-02	6.47E-02	4.15E-01	9.11E-01	6.85E-02
	LATE.DEATH	1.35E-03	8.63E-01	8.58E-01	4.24E-01	1.26E-03	9.31E-03
	SURVIVORS	2.00E-09	8.63E-01	9.55E-01	1.00E+00	2.00E-09	1.87E-07
9	EARLY.DEATH	2.75E-01	6.35E-01	9.28E-01	8.65E-03	2.68E-01	5.29E-02
	LATE.DEATH	8.58E-01	7.33E-01	5.82E-01	8.77E-01	8.58E-01	9.17E-01
	SURVIVORS	4.38E-04	9.05E-01	9.17E-01	5.55E-01	4.38E-04	7.27E-03
14	EARLY.DEATH	9.17E-01	3.86E-01	4.01E-01	1.30E-01	9.17E-01	3.73E-01
	LATE.DEATH	1.08E-01	9.90E-01	2.78E-01	7.96E-01	1.07E-01	2.91E-01
	SURVIVORS	4.38E-04	5.51E-01	9.17E-01	3.42E-01	4.38E-04	5.86E-03
15	EARLY.DEATH	2.83E-01	5.51E-01	1.34E-01	1.33E-01	2.75E-01	1.20E-01
	LATE.DEATH	1.00E+00	1.00E+00	9.47E-01	9.53E-01	1.00E+00	1.00E+00
	SURVIVORS	2.91E-01	2.44E-01	7.31E-01	1.06E-01	2.91E-01	2.78E-01
16	EARLY.DEATH	6.58E-03	1.20E-01	6.86E-01	3.73E-01	6.76E-03	7.56E-03
	LATE.DEATH	2.95E-01	5.51E-01	1.54E-01	6.93E-01	2.91E-01	2.83E-01
	SURVIVORS	9.76E-01	3.67E-01	4.50E-01	5.66E-01	9.76E-01	1.77E-01
17	EARLY.DEATH	1.31E-02	5.44E-01	5.51E-01	5.49E-01	9.58E-03	7.71E-02
	LATE.DEATH	3.30E-02	9.53E-01	8.30E-01	2.35E-02	3.16E-02	8.59E-03
	SURVIVORS	7.52E-02	5.48E-01	1.65E-01	6.68E-01	7.49E-02	1.20E-01
18	EARLY.DEATH	7.71E-02	9.35E-01	2.92E-01	7.90E-01	7.36E-02	1.20E-01
	LATE.DEATH	6.35E-01	7.35E-01	9.11E-01	7.67E-01	6.35E-01	8.67E-01
	SURVIVORS	9.17E-01	7.19E-01	4.23E-01	3.62E-01	9.17E-01	6.58E-01
20	EARLY.DEATH	6.35E-01	5.34E-02	6.41E-01	5.55E-01	6.35E-01	2.78E-01
	LATE.DEATH	6.36E-02	9.59E-01	8.58E-01	7.19E-01	5.97E-02	2.78E-01
	SURVIVORS	5.04E-04	7.23E-01	2.78E-01	2.10E-01	4.92E-04	1.14E-03
25	EARLY.DEATH	2.06E-01	3.34E-01	3.98E-01	4.01E-01	2.04E-01	2.78E-01
	LATE.DEATH	1.00E+00	9.61E-01	4.24E-01	2.51E-01	1.00E+00	2.91E-01
	SURVIVORS	7.84E-02	1.77E-01	3.98E-01	8.47E-01	7.74E-02	7.61E-02
26	EARLY.DEATH	1.34E-01	9.35E-01	5.55E-01	4.50E-01	1.54E-01	2.78E-01
	LATE.DEATH	1.65E-01	5.51E-01	2.66E-01	2.29E-02	1.59E-01	1.43E-02
	SURVIVORS	3.58E-01	4.15E-01	6.56E-01	9.17E-01	3.58E-01	3.89E-01
28	EARLY.DEATH	8.30E-01	1.07E-01	8.93E-01	9.38E-01	8.30E-01	4.03E-01
	LATE.DEATH	9.17E-01	3.37E-01	2.91E-01	6.41E-01	9.17E-01	6.35E-01
	SURVIVORS	9.04E-05	9.17E-01	9.17E-01	5.48E-01	9.04E-05	1.33E-03
29	EARLY.DEATH	9.05E-01	7.27E-03	1.95E-01	5.51E-01	9.05E-01	7.08E-02
	LATE.DEATH	1.28E-02	7.19E-01	7.19E-01	7.32E-01	1.17E-02	7.84E-02
	SURVIVORS	1.18E-06	8.17E-01	9.17E-01	9.96E-01	1.18E-06	7.60E-05

Table C. 9. Simple Effects Tests: effect of time on BCR module for sepsis mortality groups – BH adjusted p values.

P values are reported for each main effect term and for the results of the ANOVA test, comparing the variance explained by the TASC model to the variance explained by the covariate only and null models. Significant p values ($p<0.05$) are recorded in red. Results are reported to three decimal places.

21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Module	
2.18 E-01	4.24 E-02	6.64 E-01	6.70 E-01	4.11 E-02	8.01 E-01	8.83 E-01	2.89 E-01	8.23 E-01	8.75 E-01	9.68 E-02	2.50 E-01	1.56 E-01	5.76 E-01	1.09 E-01	1.69 E-01	9.48 E-01	4.10 E-01	2.82 E-01	3.16 E-01	1.66 E-01	**Arterial pH**	**Linear Model**
5.16 E-01	9.49 E-01	6.03 E-01	7.90 E-01	5.95 E-01	4.55 E-01	8.96 E-01	6.04 E-02	8.61 E-01	9.79 E-01	2.72 E-01	5.60 E-01	5.77 E-01	5.26 E-01	7.56 E-01	9.93 E-01	8.71 E-01	6.61 E-01	7.67 E-01	2.03 E-01	5.14 E-01	**Cardiogenic Pulmonary Oedema**	
4.65 E-01	2.85 E-01	5.95 E-01	7.23 E-01	9.23 E-01	5.09 E-02	8.71 E-01	5.08 E-01	3.75 E-01	1.84 E-01	9.19 E-01	7.25 E-02	9.47 E-01	5.10 E-01	4.03 E-01	8.79 E-01	3.08 E-01	1.67 E-01	9.42 E-01	6.34 E-01	5.41 E-01	**Haematocrit**	
4.34 E-01	4.64 E-01	5.73 E-01	5.32 E-01	7.46 E-01	1.89 E-01	2.99 E-01	4.31 E-01	6.69 E-01	6.77 E-01	6.35 E-01	2.10 E-01	2.27 E-01	4.73 E-01	7.22 E-01	5.61 E-01	5.94 E-01	7.73 E-01	3.69 E-01	3.82 E-01	6.20 E-01	**Infection: Number of ICU Acquired infections**	
7.16 E-01	3.82 E-02	4.41 E-01	9.04 E-02	3.49 E-01	6.70 E-01	2.00 E-02	4.77 E-01	4.84 E-01	8.62 E-01	7.74 E-01	4.51 E-01	3.77 E-02	2.70 E-01	6.43 E-01	2.41 E-01	9.22 E-01	8.94 E-01	2.81 E-02	5.73 E-01	6.07 E-01	**Infection: Ventilator Associated Pneumonia**	
5.02 E-02	1.34 E-01	1.96 E-01	5.38 E-01	6.33 E-01	3.95 E-01	8.24 E-01	6.21 E-01	1.20 E-01	1.05 E-01	2.73 E-01	7.42 E-01	8.30 E-01	7.83 E-02	5.63 E-01	7.55 E-01	5.69 E-01	3.49 E-01	3.95 E-02	7.11 E-02	6.05 E-01	**Lung Infiltrates: Diffuse Bilateral**	
3.25 E-01	7.20 E-01	4.44 E-01	9.77 E-01	2.69 E-01	3.22 E-01	4.37 E-01	4.06 E-01	3.85 E-01	7.01 E-01	9.19 E-01	7.10 E-01	7.10 E-01	3.64 E-02	3.36 E-01	9.64 E-01	6.08 E-01	5.95 E-01	2.44 E-01	7.94 E-01	9.18 E-01	**Lung Infiltrates: Lobar**	
1.17 E-01	9.65 E-01	4.94 E-01	1.77 E-01	6.05 E-01	7.77 E-01	2.86 E-01	2.88 E-01	4.12 E-01	2.85 E-01	4.50 E-01	3.82 E-01	7.31 E-01	8.71 E-01	8.99 E-01	9.68 E-01	8.72 E-01	6.92 E-01	5.14 E-01	5.91 E-02	9.57 E-01	**Lung Infiltrates: Localised**	
3.32 E-01	3.64 E-01	7.56 E-01	3.08 E-01	3.42 E-01	5.47 E-01	5.77 E-01	2.12 E-01	6.69 E-01	7.28 E-01	6.22 E-01	8.16 E-01	3.37 E-01	9.53 E-01	9.25 E-01	9.25 E-01	9.27 E-01	3.30 E-01	7.11 E-01	8.62 E-01	7.69 E-01	**Haemophilus Influenzae**	
1.69 E-01	4.39 E-01	7.15 E-01	2.25 E-01	1.97 E-01	4.23 E-01	3.17 E-01	1.94 E-01	7.76 E-01	8.63 E-01	8.31 E-01	1.59 E-01	3.81 E-01	4.30 E-01	4.75 E-01	3.54 E-01	8.44 E-01	8.38 E-01	3.19 E-01	1.57 E-01	1.28 E-01	**Streptococcus Pneumonia**	
9.16 E-01	5.22 E-01	8.73 E-01	6.33 E-01	6.60 E-02	5.52 E-01	9.12 E-01	1.44 E-01	9.07 E-01	4.37 E-01	6.72 E-01	1.88 E-01	3.98 E-01	3.41 E-01	4.54 E-01	2.06 E-01	8.22 E-01	7.32 E-01	4.71 E-01	4.86 E-01	3.25 E-01	**Potassium: Highest**	
4.73 E-01	4.27 E-01	3.24 E-01	9.57 E-01	4.04 E-01	9.73 E-01	1.77 E-01	9.17 E-01	3.24 E-01	1.25 E-01	6.14 E-01	8.33 E-01	9.00 E-01	2.71 E-01	5.53 E-01	3.10 E-01	6.56 E-01	1.49 E-01	3.12 E-01	6.46 E-01	2.87 E-01	**Sodium: Highest**	
7.46 E-01	2.43 E-01	5.89 E-01	4.10 E-01	4.10 E-01	3.13 E-01	3.98 E-01	2.23 E-01	5.78 E-01	3.71 E-01	1.88 E-01	8.31 E-01	2.52 E-01	2.46 E-01	1.84 E-01	8.94 E-01	9.38 E-01	9.69 E-01	7.17 E-01	5.26 E-02	7.75 E-01	**Therapy: Corticosteroids**	
6.50 E-02	8.58 E-01	9.63 E-01	6.80 E-01	2.56 E-01	5.35 E-01	3.81 E-01	2.08 E-01	8.54 E-01	9.03 E-01	2.81 E-01	8.30 E-01	4.76 E-01	8.06 E-01	3.36 E-01	7.84 E-01	5.80 E-01	6.39 E-01	6.92 E-01	3.41 E-01	9.35 E-01	**Therapy: Days of Inotrope Pressor Support**	
7.17 E-01	6.60 E-01	7.76 E-01	2.82 E-01	2.25 E-01	8.05 E-02	1.57 E-01	3.40 E-02	5.71 E-01	1.75 E-01	5.97 E-01	7.11 E-01	6.33 E-01	9.99 E-01	7.41 E-01	6.08 E-01	5.32 E-01	5.27 E-01	8.87 E-01	4.62 E-01	6.82 E-01	**Therapy: Drotecogin Alfa**	
4.22 E-01	2.09 E-01	5.43 E-01	1.76 E-01	9.75 E-01	5.82 E-01	6.99 E-01	1.47 E-01	8.89 E-01	7.52 E-01	6.78 E-01	3.37 E-01	4.34 E-01	5.85 E-01	9.60 E-01	7.75 E-01	6.77 E-01	1.42 E-01	2.37 E-01	1.38 E-01	5.39 E-01	**Therapy: Mechanical Ventilation CPAP**	
7.41 E-01	7.00 E-01	4.20 E-01	4.23 E-02	2.59 E-01	4.60 E-01	3.66 E-02	2.71 E-02	4.59 E-01	7.58 E-01	1.27 E-01	5.83 E-01	6.00 E-02	1.62 E-01	1.96 E-02	3.22 E-01	9.27 E-01	6.76 E-01	8.11 E-01	5.21 E-02	4.10 E-01	**Therapy: Number of Sepsis Specific Therapies**	
5.03 E-01	4.39 E-01	4.95 E-01	8.74 E-01	7.33 E-01	3.73 E-01	6.50 E-01	6.39 E-02	4.34 E-01	2.40 E-01	2.73 E-01	3.12 E-01	7.57 E-01	9.92 E-01	2.76 E-01	6.98 E-01	4.13 E-01	5.20 E-01	5.41 E-01	8.75 E-01	3.80 E-01	**Therapy: Renal Replacement Therapy**	
5.84 E-01	6.40 E-01	9.86 E-01	9.27 E-01	8.85 E-02	6.11 E-01	6.80 E-01	1.27 E-03	8.91 E-01	4.64 E-01	4.72 E-01	3.37 E-01	4.57 E-01	6.87 E-01	2.43 E-01	9.83 E-01	9.00 E-01	2.34 E-01	9.81 E-01	4.35 E-01	6.91 E-01	**Total: Apache**	
8.76 E-01	2.19 E-01	1.43 E-01	3.35 E-01	8.23 E-01	3.94 E-01	9.99 E-01	4.72 E-04	1.30 E-01	4.67 E-01	9.43 E-01	2.88 E-01	3.96 E-01	5.21 E-01	2.67 E-02	5.28 E-01	2.61 E-01	8.44 E-01	5.66 E-01	9.05 E-01	1.34 E-01	**Total: SAPS**	

Module	Linear Model											
	Arterial pH	Cardiogenic Pulmonary Oedema	Haematocrit	Infection: Number of ICU Acquired infections	Infection: Ventilator Associated Pneumonia	Lung Infiltrates: Diffuse Bilateral	Lung Infiltrates: Lobar	Lung Infiltrates: Localised	Haemophilus Influenzae	Streptococcus Pneumonia	Potassium: Highest	Sodium: Highest	Therapy: Corticosteroids	Therapy: Days of Inotrope Pressor Support	Therapy: Drotecogin Alfa	Therapy: Mechanical Ventilation CPAP	Therapy: Number of Sepsis Specific Therapies	Therapy: Renal Replacement Therapy	Total: Apache	Total: SAPS
22	3.44 E-01	1.17 E-01	2.07 E-01	6.20 E-01	2.94 E-01	9.34 E-03	8.00 E-02	3.86 E-01	9.36 E-01	8.41 E-01	9.44 E-01	3.99 E-01	3.71 E-01	9.42 E-01	3.56 E-01	1.13 E-01	7.15 E-01	2.84 E-01	8.56 E-01	3.13 E-01
23	2.89 E-01	6.28 E-01	5.44 E-02	8.52 E-01	4.65 E-01	5.27 E-01	2.70 E-01	5.06 E-01	9.39 E-01	6.28 E-01	5.52 E-01	6.28 E-01	4.65 E-01	8.65 E-01	4.47 E-01	6.43 E-01	4.10 E-01	4.88 E-01	8.92 E-01	4.94 E-02
24	2.59 E-01	2.32 E-01	7.48 E-01	6.84 E-01	2.67 E-01	8.86 E-01	9.08 E-01	8.89 E-01	7.08 E-01	2.74 E-01	9.70 E-01	3.08 E-01	6.20 E-01	9.98 E-01	3.46 E-01	9.90 E-01	4.96 E-01	4.39 E-01	4.15 E-01	4.41 E-01
25	6.51 E-01	7.83 E-01	6.12 E-01	7.05 E-01	5.46 E-01	9.02 E-01	1.90 E-01	4.98 E-01	7.19 E-01	9.96 E-01	7.77 E-01	9.66 E-01	6.72 E-01	3.75 E-01	6.09 E-01	5.75 E-01	3.20 E-01	4.08 E-02	3.67 E-01	1.43 E-01
26	4.63 E-01	1.04 E-01	9.80 E-01	3.59 E-01	7.11 E-01	1.43 E-01	3.97 E-01	1.51 E-01	3.93 E-01	8.18 E-01	4.30 E-01	8.95 E-01	1.07 E-01	3.46 E-01	7.82 E-01	8.09 E-01	8.57 E-01	8.93 E-01	7.56 E-01	9.85 E-01
27	3.19 E-01	3.79 E-01	7.18 E-01	4.66 E-01	2.44 E-01	9.81 E-01	9.80 E-01	8.89 E-01	8.27 E-01	8.87 E-01	9.84 E-01	5.34 E-01	1.24 E-01	7.09 E-01	9.72 E-01	5.60 E-01	5.23 E-01	6.43 E-01	9.41 E-01	6.16 E-01
28	1.25 E-01	3.36 E-03	9.08 E-01	5.89 E-01	1.22 E-01	2.99 E-01	9.97 E-01	2.84 E-01	8.06 E-01	6.31 E-02	4.83 E-01	9.84 E-01	4.31 E-01	4.02 E-01	8.80 E-01	1.80 E-01	5.57 E-01	9.57 E-01	3.93 E-01	3.10 E-01
29	1.61 E-01	4.55 E-01	5.98 E-01	2.70 E-01	3.18 E-02	1.64 E-01	5.34 E-01	6.49 E-01	8.39 E-01	3.57 E-01	5.94 E-01	8.82 E-01	9.11 E-01	7.26 E-01	9.16 E-01	1.08 E-01	6.13 E-01	6.32 E-01	9.18 E-01	4.35 E-01
30	1.21 E-01	8.71 E-01	4.83 E-01	1.59 E-01	4.55 E-02	6.94 E-01	4.59 E-01	8.51 E-01	2.75 E-01	3.20 E-01	2.38 E-01	4.12 E-01	7.40 E-01	2.02 E-01	4.62 E-01	6.10 E-04	9.28 E-01	1.94 E-01	5.57 E-01	1.07 E-01
31	4.00 E-01	8.76 E-01	8.58 E-01	4.43 E-01	5.95 E-01	6.08 E-01	5.36 E-01	8.23 E-01	8.93 E-01	9.26 E-01	7.44 E-01	6.20 E-02	5.03 E-01	3.79 E-01	8.00 E-01	5.39 E-01	3.19 E-01	6.37 E-01	7.82 E-01	2.28 E-01

Table C. 10. Linear models comparing the effect of clinical variables measured once per ICU admission, on BCR module score measured once using day 1 samples.

Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Module	
2.11 E-01	8.52 E-02	2.68 E-02	4.66 E-01	1.63 E-01	4.81 E-01	5.75 E-01	2.47 E-01	1.04 E-01	4.68 E-01	1.93 E-01	1.43 E-01	7.85 E-02	1.65 E-01	7.77 E-01	6.41 E-01	9.55 E-01	7.46 E-01	5.07 E-02	9.44 E-03	5.31 E-01	7.26 E-01	Bicarbonate	Linear Mixed Effect Model
5.72 E-02	1.94 E-03	4.58 E-06	9.68 E-01	7.09 E-02	2.84 E-01	4.44 E-03	6.87 E-01	1.15 E-05	7.13 E-01	3.32 E-01	1.38 E-01	4.45 E-02	1.61 E-02	7.12 E-05	1.54 E-01	9.79 E-02	1.54 E-02	9.98 E-01	4.49 E-14	2.54 E-01	2.72 E-03	Bilirubin: Highest	
3.17 E-01	2.12 E-03	5.25 E-01	1.04 E-01	3.57 E-01	6.96 E-01	5.06 E-01	3.54 E-01	4.77 E-01	1.48 E-01	1.82 E-01	9.37 E-01	2.52 E-01	8.03 E-01	1.20 E-01	6.97 E-01	3.54 E-02	3.80 E-01	2.33 E-01	5.49 E-01	4.20 E-01	2.49 E-01	Creatine: Highest	
4.63 E-02	2.47 E-03	1.24 E-04	6.25 E-01	2.97 E-01	3.37 E-01	9.71 E-03	5.73 E-01	9.95 E-06	4.41 E-01	5.62 E-01	6.39 E-01	1.63 E-01	4.97 E-02	2.80 E-05	1.26 E-01	1.93 E-02	2.59 E-02	7.72 E-01	1.78 E-12	2.29 E-01	2.88 E-03	GCS: Total	
9.79 E-01	7.44 E-01	2.91 E-01	2.68 E-01	5.22 E-01	7.38 E-01	9.10 E-01	7.73 E-01	4.81 E-01	4.23 E-01	7.30 E-01	3.09 E-02	1.22 E-01	8.17 E-01	6.92 E-01	4.12 E-01	8.92 E-01	3.77 E-01	9.18 E-01	3.49 E-01	5.02 E-01	5.09 E-01	Heart Rate: Highest	
1.26 E-02	8.16 E-03	1.31 E-04	9.78 E-01	2.31 E-01	2.69 E-01	8.63 E-03	7.01 E-01	7.57 E-06	6.61 E-01	6.56 E-01	9.16 E-01	6.99 E-02	3.65 E-02	2.48 E-05	1.76 E-01	1.68 E-02	1.10 E-02	7.92 E-01	2.96 E-13	2.16 E-01	3.50 E-03	Hypertension	
2.85 E-01	2.37 E-01	3.19 E-01	1.08 E-01	1.63 E-01	2.57 E-01	1.97 E-01	3.45 E-01	3.45 E-01	1.44 E-01	2.20 E-01	6.14 E-01	4.11 E-01	5.34 E-01	6.18 E-01	3.97 E-01	1.71 E-01	5.06 E-01	9.49 E-01	5.94 E-01	8.72 E-01	4.04 E-02	Leukocytes: Lymphocytes	
1.67 E-02	1.31 E-02	4.40 E-05	9.88 E-01	2.15 E-01	2.98 E-01	5.27 E-03	5.68 E-01	5.03 E-06	6.64 E-01	7.36 E-01	5.66 E-01	1.55 E-01	2.57 E-02	2.15 E-05	1.81 E-01	8.92 E-03	1.35 E-02	7.67 E-01	1.29 E-13	1.77 E-01	8.98 E-03	Leukocytes: Mononuclear	
2.57 E-01	9.46 E-01	9.53 E-01	1.90 E-01	1.00 E-01	9.71 E-01	3.00 E-01	6.22 E-01	9.27 E-01	2.97 E-01	4.54 E-01	1.39 E-01	8.99 E-03	7.31 E-01	8.66 E-01	1.57 E-02	9.94 E-01	8.92 E-01	2.76 E-01	3.44 E-01	5.73 E-01	5.65 E-01	Leukocytes: Polymorphonuclear	
5.94 E-03	1.23 E-02	9.48 E-05	5.03 E-01	1.21 E-01	2.17 E-01	6.36 E-03	5.15 E-01	2.55 E-05	3.29 E-01	4.67 E-01	9.46 E-01	2.47 E-02	2.82 E-02	6.04 E-05	1.84 E-02	2.29 E-02	2.07 E-02	9.84 E-01	2.58 E-13	1.30 E-01	3.83 E-03	Mean Arterial Pressure: Lowest	
2.63 E-01	3.32 E-01	3.44 E-01	8.74 E-01	3.30 E-01	8.80 E-01	8.61 E-01	1.32 E-01	1.78 E-01	9.82 E-01	1.92 E-01	6.51 E-01	3.79 E-01	5.57 E-01	4.30 E-01	3.81 E-01	5.81 E-04	8.17 E-01	9.02 E-01	8.63 E-01	2.47 E-01	7.47 E-01	MPa02	
2.01 E-02	6.19 E-03	1.17 E-04	8.74 E-01	2.10 E-01	2.54 E-01	8.20 E-03	8.66 E-01	3.67 E-06	5.50 E-01	4.50 E-01	6.76 E-01	9.05 E-02	4.06 E-02	9.04 E-06	1.76 E-01	4.53 E-02	1.99 E-02	7.58 E-01	1.28 E-13	2.45 E-01	4.29 E-03	PC02	
3.58 E-01	1.84 E-01	7.78 E-01	5.60 E-01	7.02 E-01	7.22 E-01	4.92 E-01	7.13 E-01	2.33 E-02	8.36 E-01	4.11 E-01	5.97 E-01	7.21 E-01	9.01 E-01	2.29 E-01	4.18 E-01	2.85 E-01	3.66 E-01	4.18 E-01	3.49 E-01	9.14 E-01	6.40 E-02	Platelets: Lowest	
5.03 E-02	6.02 E-03	2.91 E-04	7.81 E-01	3.86 E-01	4.01 E-01	1.04 E-02	6.00 E-01	3.15 E-06	5.40 E-01	9.09 E-01	4.95 E-01	1.44 E-01	4.50 E-02	3.20 E-04	1.85 E-01	7.13 E-02	1.35 E-02	7.51 E-01	1.79 E-11	2.66 E-01	3.25 E-03	Respiratory Rate	
7.80 E-01	2.96 E-01	8.74 E-01	3.23 E-01	5.87 E-01	5.91 E-01	4.45 E-01	4.88 E-01	7.29 E-01	4.28 E-01	3.39 E-01	3.41 E-01	8.41 E-01	5.85 E-01	7.41 E-01	9.49 E-01	5.59 E-01	8.86 E-01	3.92 E-01	9.89 E-01	5.83 E-01	1.93 E-01	Systolic Blood Pressure: Lowest	
2.26 E-02	1.98 E-02	2.02 E-04	8.94 E-01	4.09 E-01	6.53 E-02	3.20 E-02	5.47 E-01	2.66 E-05	5.68 E-01	7.44 E-01	5.84 E-01	1.31 E-01	4.00 E-02	1.67 E-04	5.86 E-02	1.11 E-01	2.08 E-02	5.89 E-01	1.45 E-11	1.71 E-01	1.70 E-02	Temperature: Highest	
1.26 E-01	5.57 E-01	9.75 E-01	7.05 E-01	6.69 E-01	5.44 E-01	5.12 E-01	3.24 E-01	9.98 E-01	3.78 E-01	2.31 E-01	1.73 E-01	5.52 E-04	4.79 E-01	7.07 E-02	2.14 E-01	3.52 E-01	7.16 E-01	3.64 E-01	5.91 E-01	2.11 E-01	6.36 E-01	Total: SOFA	
1.47 E-02	8.61 E-03	2.54 E-04	7.81 E-01	3.62 E-01	3.77 E-01	1.56 E-02	8.06 E-01	2.89 E-05	4.40 E-01	8.14 E-01	4.76 E-01	1.18 E-01	4.83 E-02	8.38 E-05	1.41 E-01	9.58 E-02	2.43 E-02	6.35 E-01	1.63 E-11	1.91 E-01	8.19 E-03	Urea: Highest	
9.22 E-01	6.71 E-01	1.52 E-01	7.28 E-01	1.67 E-02	4.92 E-01	7.69 E-01	3.25 E-01	2.96 E-02	6.60 E-01	6.64 E-01	9.21 E-02	1.07 E-01	5.67 E-01	9.24 E-01	1.01 E-01	1.75 E-01	7.95 E-01	4.92 E-01	7.81 E-01	5.89 E-01	6.12 E-01	WCC: Highest	

Module	Linear Mixed Effect Model																					
	Bicarbonate		Bilirubin: Highest		Creatine: Highest		GCS: Total		Heart Rate: Highest		Hypertension		Leukocytes: Lymphocytes		Leukocytes: Monomuclear		Leukocytes: Polymorphonuclear		Mean Arterial Pressure: Lowest		MPa02	
23	1.37	E-01	2.60	E-01	3.48	E-01	9.66	E-02	5.57	E-01	2.27	E-01	3.53	E-01	2.15	E-01	2.77	E-01	8.78	E-02	7.83	E-01
24	1.73	E-02	9.23	E-01	4.20	E-01	4.37	E-01	6.50	E-01	6.40	E-01	1.29	E-01	6.93	E-01	7.28	E-01	4.77	E-01	8.35	E-01
25	5.74	E-02	1.30	E-01	2.17	E-01	2.57	E-01	7.05	E-01	2.40	E-01	3.17	E-01	2.06	E-01	4.78	E-01	3.69	E-01	5.48	E-01
26	3.69	E-01	2.41	E-02	4.79	E-01	1.31	E-02	4.54	E-01	4.99	E-03	8.96	E-01	7.96	E-03	4.99	E-01	6.35	E-03	5.90	E-01
27	8.02	E-01	1.18	E-02	7.24	E-01	1.50	E-02	9.90	E-01	1.98	E-02	7.82	E-01	1.72	E-02	8.10	E-01	2.54	E-02	7.30	E-01
28	1.49	E-01	3.97	E-04	9.40	E-01	9.30	E-04	3.17	E-01	1.02	E-03	7.33	E-01	5.69	E-04	5.62	E-02	7.08	E-05	4.23	E-01
29	3.14	E-02	3.49	E-09	5.01	E-01	2.53	E-08	2.66	E-02	1.77	E-08	5.52	E-01	3.73	E-09	3.25	E-01	6.55	E-09	5.94	E-01
30	4.27	E-02	1.63	E-01	7.96	E-01	7.63	E-01	1.67	E-01	5.10	E-01	3.80	E-01	6.29	E-01	1.10	E-02	2.72	E-01	5.73	E-01
31	8.99	E-01	6.22	E-02	4.27	E-01	5.20	E-02	4.33	E-01	3.41	E-02	5.19	E-01	3.89	E-02	2.85	E-01	1.22	E-01	6.07	E-01

Module	Platelets: Lowest		Respiratory Rate		Systolic Blood Pressure: Lowest		Temperature: Highest		Total: SOFA		Urea: Highest		WCC: Highest	
23	9.10	E-01	1.21	E-01	8.06	E-01	9.99	E-02	1.55	E-01	6.57	E-02	9.04	E-01
24	8.27	E-01	5.24	E-01	4.87	E-01	4.41	E-01	2.70	E-01	3.70	E-01	9.56	E-01
25	4.84	E-01	7.45	E-01	9.91	E-01	6.17	E-01	5.89	E-01	6.56	E-01	6.08	E-01
26	7.50	E-01	1.31	E-02	4.13	E-01	2.11	E-02	8.92	E-01	1.47	E-02	7.42	E-01
27	1.47	E-01	1.74	E-02	3.09	E-01	3.14	E-02	5.74	E-01	4.32	E-02	2.65	E-01
28	3.09	E-01	1.20	E-03	4.81	E-01	1.95	E-03	7.31	E-01	2.39	E-03	2.39	E-01
29	6.90	E-01	4.04	E-07	9.28	E-01	2.56	E-07	7.71	E-01	2.89	E-07	8.44	E-01
30	8.00	E-01	8.51	E-01	3.22	E-01	9.92	E-01	9.66	E-01	9.23	E-01	6.87	E-01
31	6.21	E-01	5.08	E-02	3.95	E-01	3.65	E-02	8.25	E-01	4.06	E-02	1.13	E-01

Table C. 11. Linear models comparing the effect of longitudinal clinical variables on BCR module score using all samples.

Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

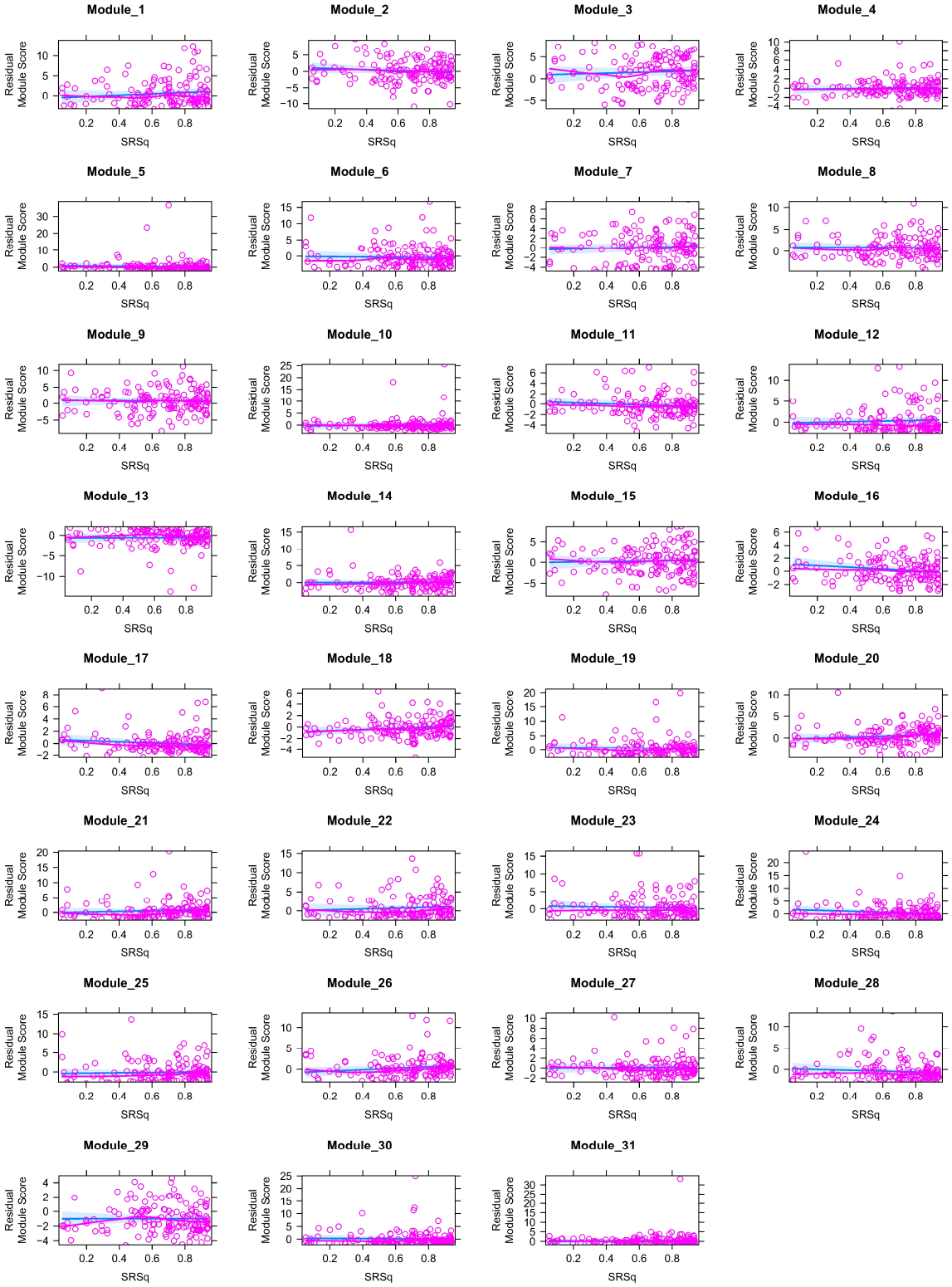


Figure C. 2. Partial residual plots showing the association between SRSq and BCR module score controlling for the covariates age, sex, CI and time point.

SRSq was significantly associated with module 26 score.

Module	Linear Mixed Effect Model					ANOVA	
	TIME	SRSq	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values							
1	7.36E-04	2.16E-01	8.44E-03	1.73E-01	1.16E-02	3.05E-03	6.84E-04
2	5.96E-01	1.97E-01	8.20E-01	1.15E-01	7.74E-01	1.77E-01	2.99E-01
3	7.37E-12	4.96E-01	4.30E-02	3.10E-01	1.95E-01	1.74E-11	8.92E-11
4	8.62E-01	5.58E-01	1.20E-01	4.01E-01	4.90E-02	7.59E-01	3.30E-01
5	1.96E-02	5.83E-01	2.45E-02	4.52E-01	2.92E-01	5.94E-02	5.23E-02
6	2.71E-02	7.88E-01	4.48E-01	5.12E-01	2.74E-01	5.48E-02	1.88E-01
7	7.35E-01	2.03E-01	8.10E-01	2.89E-01	9.21E-01	2.13E-01	5.06E-01
8	3.05E-05	9.07E-01	9.84E-01	9.91E-01	5.69E-01	1.56E-05	4.01E-04
9	7.59E-02	9.56E-01	4.80E-01	8.86E-01	3.26E-02	1.26E-01	9.37E-02
10	1.08E-01	5.49E-01	1.34E-01	6.72E-01	2.24E-01	2.73E-01	3.63E-01
11	3.24E-01	1.61E-01	4.96E-01	2.29E-01	8.75E-01	3.59E-01	5.42E-01
12	3.30E-01	4.27E-01	2.43E-01	8.71E-01	2.72E-01	5.85E-01	7.24E-01
13	6.38E-01	8.95E-01	3.85E-03	1.77E-01	1.54E-01	8.32E-01	7.39E-02
14	8.12E-05	9.36E-01	6.84E-01	4.20E-01	9.08E-01	1.27E-04	2.38E-03
15	8.04E-01	6.55E-01	6.92E-01	4.80E-01	6.15E-02	9.04E-01	4.03E-01
16	8.59E-04	1.09E-01	3.85E-03	7.07E-01	5.31E-01	3.55E-03	7.64E-04
17	1.41E-01	1.92E-01	3.00E-01	1.55E-01	3.26E-01	2.67E-01	3.06E-01
18	7.47E-02	1.25E-01	9.40E-01	9.15E-01	5.46E-01	1.41E-01	4.67E-01
19	9.96E-01	7.15E-01	5.06E-03	1.55E-01	1.20E-01	9.20E-01	9.69E-02
20	3.54E-05	2.07E-01	1.74E-01	6.04E-01	9.14E-01	1.85E-04	1.12E-03
21	3.03E-03	3.29E-01	5.46E-03	1.33E-01	1.47E-01	1.16E-02	2.13E-03
22	7.33E-03	3.93E-01	1.77E-01	9.76E-01	5.71E-01	2.50E-02	9.27E-02
23	4.38E-01	6.20E-01	2.58E-01	7.58E-01	3.60E-01	4.91E-01	7.11E-01
24	9.60E-01	1.81E-01	3.30E-01	3.71E-01	8.43E-01	3.42E-01	4.70E-01
25	4.20E-01	5.99E-01	4.12E-01	9.47E-01	7.74E-01	4.00E-01	6.40E-01
26	4.33E-04	4.82E-02	4.53E-01	4.37E-01	1.16E-01	2.56E-03	8.75E-03
27	3.92E-02	8.50E-01	2.65E-01	1.30E-01	2.86E-01	8.13E-02	1.20E-01
28	1.12E-05	2.88E-01	4.00E-01	5.72E-01	6.73E-01	5.22E-05	7.86E-04
29	3.65E-08	9.79E-01	3.13E-02	4.44E-01	3.15E-01	1.56E-08	6.87E-08
30	5.38E-01	8.43E-01	5.75E-01	1.20E-01	7.84E-01	8.26E-01	5.96E-01
31	1.12E-01	5.74E-01	1.53E-02	2.60E-01	8.69E-02	1.16E-01	5.43E-02
BH Adjusted p values							
1	8.53E-03	4.83E-01	5.39E-02	4.37E-01	6.82E-02	2.46E-02	8.53E-03
2	7.79E-01	4.60E-01	9.37E-01	3.68E-01	9.08E-01	4.37E-01	5.76E-01
3	1.60E-09	7.22E-01	2.03E-01	5.85E-01	4.60E-01	1.89E-09	6.46E-09
4	9.54E-01	7.70E-01	3.68E-01	6.62E-01	2.21E-01	9.01E-01	5.87E-01
5	1.09E-01	7.74E-01	1.32E-01	6.92E-01	5.72E-01	2.48E-01	2.31E-01
6	1.40E-01	9.15E-01	6.92E-01	7.36E-01	5.55E-01	2.33E-01	4.52E-01
7	8.82E-01	4.69E-01	9.30E-01	5.70E-01	9.65E-01	4.82E-01	7.32E-01
8	7.35E-04	9.65E-01	9.93E-01	9.96E-01	7.70E-01	4.24E-04	5.79E-03
9	2.94E-01	9.83E-01	7.14E-01	9.65E-01	1.61E-01	3.75E-01	3.39E-01
10	3.68E-01	7.64E-01	3.83E-01	8.40E-01	4.95E-01	5.55E-01	6.25E-01
11	5.87E-01	4.21E-01	7.22E-01	5.01E-01	9.58E-01	6.24E-01	7.64E-01
12	5.87E-01	6.81E-01	5.28E-01	9.58E-01	5.55E-01	7.74E-01	8.72E-01
13	8.12E-01	9.65E-01	2.78E-02	4.37E-01	4.10E-01	9.41E-01	2.94E-01
14	1.47E-03	9.76E-01	8.48E-01	6.76E-01	9.65E-01	2.11E-03	2.15E-02
15	9.28E-01	8.26E-01	8.53E-01	7.14E-01	2.52E-01	9.65E-01	6.62E-01
16	8.88E-03	3.68E-01	2.78E-02	8.66E-01	7.59E-01	2.75E-02	8.53E-03
17	3.93E-01	4.58E-01	5.76E-01	4.10E-01	5.87E-01	5.55E-01	5.82E-01
18	2.94E-01	3.75E-01	9.76E-01	9.65E-01	7.64E-01	3.93E-01	7.08E-01
19	9.96E-01	8.67E-01	3.54E-02	4.10E-01	3.68E-01	9.65E-01	3.45E-01
20	7.69E-04	4.72E-01	4.37E-01	7.80E-01	9.65E-01	2.86E-03	1.11E-02
21	2.46E-02	5.87E-01	3.71E-02	3.83E-01	4.04E-01	6.82E-02	2.01E-02
22	4.82E-02	6.62E-01	4.37E-01	9.93E-01	7.70E-01	1.32E-01	3.39E-01
23	6.90E-01	7.96E-01	5.52E-01	9.01E-01	6.24E-01	7.22E-01	8.66E-01
24	9.83E-01	4.40E-01	5.87E-01	6.33E-01	9.43E-01	6.04E-01	7.08E-01
25	6.76E-01	7.79E-01	6.72E-01	9.78E-01	9.08E-01	6.62E-01	8.12E-01
26	5.88E-03	2.21E-01	6.92E-01	6.90E-01	3.68E-01	2.22E-02	5.43E-02
27	1.89E-01	9.45E-01	5.55E-01	3.81E-01	5.70E-01	3.09E-01	3.68E-01
28	3.48E-04	5.70E-01	6.62E-01	7.70E-01	8.40E-01	1.03E-03	8.53E-03
29	1.58E-06	9.93E-01	1.58E-01	6.92E-01	5.87E-01	8.47E-07	2.48E-06
30	7.63E-01	9.43E-01	7.70E-01	3.68E-01	9.14E-01	9.38E-01	7.79E-01
31	3.68E-01	7.70E-01	8.72E-02	5.52E-01	3.25E-01	3.68E-01	2.33E-01

Table C. 12. Linear Mixed Effects models investigating the effect of SRSq on BCR module score.

P values are reported for each main effect term and also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the null and covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

Cell Type	Linear Mixed Effect Model						ANOVA	
	DISEASE.SEPSIS	TIMEPOINT	TIMEPOINTx SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values								
B cells	8.93E-05	9.24E-01	8.82E-02	8.93E-01	6.16E-01	1.60E-01	4.01E-06	6.33E-05
CD8 T cells	1.40E-05	7.06E-01	2.72E-02	7.08E-01	4.90E-01	6.93E-01	4.54E-07	8.06E-08
cDCs	3.18E-10	5.27E-01	2.93E-02	3.76E-01	4.13E-01	8.98E-01	8.30E-15	1.45E-15
Classical monocytes	2.82E-03	8.35E-01	6.86E-02	7.45E-01	9.18E-01	1.18E-01	2.74E-04	4.09E-03
Cycling neutrophil progenitors	1.16E-03	9.23E-01	3.72E-02	2.04E-01	8.14E-01	6.90E-01	2.52E-06	2.70E-05
Cycling TNK	5.77E-01	6.89E-01	6.00E-01	4.04E-01	4.43E-01	3.98E-01	2.76E-01	4.26E-01
Degranulating neutrophils	1.20E-07	9.25E-01	2.28E-01	9.82E-01	9.93E-01	4.02E-01	1.68E-08	7.15E-12
Eosinophils	5.77E-05	2.82E-01	7.47E-01	3.38E-01	8.07E-02	3.96E-01	2.61E-07	5.45E-10
HSPCs	1.65E-01	8.55E-01	3.44E-01	2.79E-01	8.02E-01	4.71E-01	2.37E-01	4.35E-01
IL1R2 immature neutrophils	6.51E-05	1.00E+00	4.28E-04	1.65E-02	8.33E-01	4.54E-01	3.34E-13	8.40E-13
Mast cells eosinophils	1.56E-06	6.19E-01	6.67E-02	8.70E-01	2.22E-01	5.77E-01	8.47E-07	7.41E-09
Mature neutrophils	7.71E-01	7.92E-01	5.74E-01	3.58E-02	3.55E-02	2.88E-01	8.99E-01	2.06E-01
Memory CD4 T cells	6.11E-11	7.66E-02	4.71E-03	3.69E-01	9.94E-01	6.40E-01	5.02E-10	1.06E-13
MPO. immature neutrophils or progenitors	7.74E-01	9.99E-01	1.18E-03	7.43E-01	3.42E-01	9.05E-01	2.21E-11	2.57E-10
Naive CD4 T cells	1.79E-11	6.60E-01	2.53E-01	1.02E-01	2.99E-01	5.69E-02	5.15E-11	5.72E-16
Naive CD8 T cells	1.49E-19	3.19E-01	4.96E-02	9.41E-01	6.28E-02	9.05E-01	6.83E-24	6.70E-28
NK cells	1.24E-21	3.33E-02	4.78E-02	9.49E-01	8.28E-01	6.06E-01	7.57E-19	4.11E-24
Non classical monocytes	5.11E-17	7.38E-01	7.67E-01	1.89E-01	6.46E-01	7.62E-01	3.19E-18	4.78E-22
PADI4 immature neutrophils	1.08E-01	9.53E-01	2.07E-01	2.42E-01	4.33E-01	8.14E-01	8.73E-04	3.46E-03
pDCs	8.88E-02	6.55E-01	5.93E-01	1.74E-01	1.25E-02	2.87E-01	3.00E-02	8.34E-03
Plasmablasts	8.10E-01	5.66E-01	1.34E-01	6.76E-01	4.06E-01	5.13E-01	1.81E-04	2.37E-03
Platelets	7.73E-02	7.37E-01	7.96E-01	2.68E-01	2.16E-01	1.12E-01	2.67E-01	3.35E-01
S100A8 9 hi neutrophils	1.51E-06	7.98E-01	3.72E-01	5.84E-01	2.42E-01	8.65E-01	9.29E-06	8.85E-08
BH Adjusted p values								
B cells	3.82E-04	9.67E-01	2.18E-01	9.67E-01	8.44E-01	3.63E-01	2.05E-05	2.84E-04
CD8 T cells	6.76E-05	8.98E-01	8.64E-02	8.98E-01	7.34E-01	8.91E-01	2.69E-06	5.49E-07
cDCs	2.66E-09	7.76E-01	9.13E-02	6.29E-01	6.55E-01	9.67E-01	1.27E-13	2.43E-14
Classical monocytes	1.02E-02	9.37E-01	1.80E-01	9.17E-01	9.67E-01	2.76E-01	1.12E-03	1.42E-02
Cycling neutrophil progenitors	4.42E-03	9.67E-01	1.07E-01	4.38E-01	9.31E-01	8.91E-01	1.33E-05	1.28E-04
Cycling TNK	8.23E-01	8.91E-01	8.37E-01	6.50E-01	6.79E-01	6.50E-01	5.24E-01	6.70E-01
Degranulating neutrophils	7.61E-07	9.67E-01	4.66E-01	1.00E+00	1.00E+00	6.50E-01	1.19E-07	8.22E-11
Eosinophils	2.66E-04	5.24E-01	9.17E-01	5.91E-01	2.03E-01	6.50E-01	1.60E-06	4.18E-09
HSPCs	3.71E-01	9.53E-01	5.91E-01	5.24E-01	9.31E-01	7.10E-01	4.79E-01	6.72E-01
IL1R2 immature neutrophils	2.85E-04	1.00E+00	1.71E-03	5.31E-02	9.37E-01	6.91E-01	4.38E-12	1.03E-11
Mast cells eosinophils	8.46E-06	8.44E-01	1.78E-01	9.59E-01	4.60E-01	8.23E-01	4.87E-06	5.45E-08
Mature neutrophils	9.25E-01	9.31E-01	8.23E-01	1.05E-01	1.05E-01	5.24E-01	9.67E-01	4.38E-01
Memory CD4 T cells	5.63E-10	1.97E-01	1.60E-02	6.29E-01	1.00E+00	8.66E-01	4.02E-09	1.50E-12
MPO. immature neutrophils or progenitors	9.25E-01	1.00E+00	4.42E-03	9.17E-01	5.91E-01	9.67E-01	2.26E-10	2.25E-09
Naive CD4 T cells	1.94E-10	8.74E-01	4.95E-01	2.47E-01	5.39E-01	1.56E-01	4.99E-10	1.05E-14
Naive CD8 T cells	4.58E-18	5.71E-01	1.38E-01	9.78E-01	1.70E-01	9.67E-01	4.19E-22	1.23E-25
NK cells	4.58E-20	1.00E-01	1.35E-01	9.80E-01	9.37E-01	8.38E-01	1.99E-17	3.78E-22
Non classical monocytes	1.04E-15	9.17E-01	9.25E-01	4.14E-01	8.68E-01	9.25E-01	7.33E-17	2.20E-20
PADI4 immature neutrophils	2.59E-01	9.80E-01	4.38E-01	4.79E-01	6.72E-01	9.31E-01	3.42E-03	1.22E-02
pDCs	2.18E-01	8.73E-01	8.33E-01	3.87E-01	4.09E-02	5.24E-01	9.19E-02	2.79E-02
Plasmablasts	9.31E-01	8.23E-01	3.08E-01	8.89E-01	6.50E-01	7.62E-01	7.59E-04	8.70E-03
Platelets	1.97E-01	9.17E-01	9.31E-01	5.13E-01	4.51E-01	2.65E-01	5.13E-01	5.91E-01
S100A8 9 hi neutrophils	8.41E-06	9.31E-01	6.29E-01	8.26E-01	4.79E-01	9.59E-01	4.62E-05	5.81E-07

Table C. 13. Linear mixed effects models investigating the effect of disease status and time on CIBERSORTx estimated cell fractions.

P values are reported for each main effect term and also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the null and covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to three decimal places.

Cell Type	Day	Simple Effects Tests: Health versus Sepsis				ANOVA
		DISEASE SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL
Raw p values						
CD8 T cells	Day1	2.62E-05	4.87E-01	4.56E-01	5.85E-01	2.62E-05
	Day3	7.78E-04	6.38E-01	6.36E-01	7.99E-01	7.78E-04
	Day 5	9.80E-10	3.22E-03	6.15E-01	1.21E-01	9.80E-10
cDCs	Day1	5.53E-09	1.55E-01	1.86E-01	9.05E-01	5.53E-09
	Day3	1.14E-05	1.49E-01	5.66E-01	2.26E-01	1.14E-05
	Day 5	5.48E-04	2.66E-01	9.03E-01	3.31E-01	5.48E-04
Cycling neutrophil progenitors	Day1	4.37E-03	2.67E-01	4.77E-01	5.09E-01	4.37E-03
	Day3	3.47E-01	4.10E-02	3.15E-01	7.36E-02	3.47E-01
	Day 5	5.46E-01	8.63E-01	6.37E-01	9.46E-01	5.46E-01
IL1R2 immature neutrophils	Day1	1.17E-02	1.86E-02	9.59E-01	2.80E-01	1.17E-02
	Day3	2.50E-01	5.01E-02	9.94E-01	1.51E-01	2.50E-01
	Day 5	3.08E-01	9.15E-01	1.57E-01	9.75E-01	3.08E-01
Memory CD4 T cells	Day1	3.26E-14	9.11E-01	4.65E-01	3.91E-01	3.26E-14
	Day3	8.98E-03	4.31E-01	7.83E-01	5.21E-01	8.98E-03
	Day 5	6.75E-07	4.74E-03	8.87E-01	3.66E-01	6.75E-07
MPO immature neutrophils or progenitors	Day1	4.69E-02	7.96E-01	5.01E-02	5.39E-01	4.69E-02
	Day3	3.38E-02	9.30E-01	4.04E-01	9.10E-01	3.38E-02
	Day 5	2.94E-02	5.63E-01	6.63E-01	8.18E-01	2.94E-02
Naïve CD8 T cells	Day1	6.54E-17	2.14E-01	2.25E-01	5.16E-01	6.54E-17
	Day3	1.10E-14	4.10E-01	6.72E-02	4.80E-01	1.10E-14
	Day 5	6.24E-10	1.14E-02	1.93E-01	1.56E-01	6.24E-10
NK cells	Day1	5.68E-14	9.76E-01	6.38E-01	4.93E-01	5.68E-14
	Day3	1.53E-16	9.79E-01	6.34E-01	8.92E-01	1.53E-16
	Day 5	4.42E-13	3.09E-01	4.45E-02	7.55E-01	4.42E-13
BH Adjusted p values						
CD8 T cells	Day1	1.31E-04	7.04E-01	6.93E-01	7.55E-01	1.31E-04
	Day3	3.33E-03	7.74E-01	7.74E-01	9.22E-01	3.33E-03
	Day 5	7.35E-09	1.33E-02	7.74E-01	2.85E-01	7.35E-09
CD8 T cells	Day1	3.69E-08	3.35E-01	3.92E-01	9.71E-01	3.69E-08
	Day3	6.20E-05	3.35E-01	7.39E-01	4.45E-01	6.20E-05
	Day 5	2.53E-03	4.93E-01	9.71E-01	5.60E-01	2.53E-03
Cycling neutrophil progenitors	Day1	1.69E-02	4.93E-01	7.03E-01	7.18E-01	1.69E-02
	Day3	5.70E-01	1.15E-01	5.41E-01	1.77E-01	5.70E-01
	Day 5	7.27E-01	9.71E-01	7.74E-01	9.87E-01	7.27E-01
IL1R2 immature neutrophils	Day1	3.78E-02	5.89E-02	9.87E-01	5.09E-01	3.78E-02
	Day3	4.76E-01	1.25E-01	9.94E-01	3.35E-01	4.76E-01
	Day 5	5.38E-01	9.71E-01	3.35E-01	9.87E-01	5.38E-01
Memory CD4 T cells	Day1	4.90E-13	9.71E-01	6.97E-01	6.25E-01	4.90E-13
	Day3	3.17E-02	6.63E-01	9.21E-01	7.18E-01	3.17E-02
	Day 5	4.05E-06	1.78E-02	9.71E-01	5.93E-01	4.05E-06
MPO immature neutrophils or progenitors	Day1	1.22E-01	9.22E-01	1.25E-01	7.27E-01	1.22E-01
	Day3	9.66E-02	9.79E-01	6.38E-01	9.71E-01	9.66E-02
	Day 5	8.83E-02	7.39E-01	7.96E-01	9.35E-01	8.83E-02
Naïve CD8 T cells	Day1	3.92E-15	4.35E-01	4.45E-01	7.18E-01	3.92E-15
	Day3	2.21E-13	6.39E-01	1.64E-01	7.03E-01	2.21E-13
	Day 5	5.35E-09	3.78E-02	3.99E-01	3.35E-01	5.35E-09
NK cells	Day1	6.82E-13	9.87E-01	7.74E-01	7.04E-01	6.82E-13
	Day3	4.60E-15	9.87E-01	7.74E-01	9.71E-01	4.60E-15
	Day 5	4.42E-12	5.38E-01	1.21E-01	8.96E-01	4.42E-12

Table C. 14. Simple effects tests comparing CIBERSORTx estimate cell fractions between health and sepsis at each time point.

Raw p values are reported for each main effect term. Raw p values are also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

Group	Cell Type	Simple Effects Tests: Effect of Time on health/sepsis				ANOVA	
		TIME POINT	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values							
Health	**CD8 T cells**	6.65E-01	2.61E-01	2.76E-01	0.00E+00	6.62E-01	3.37E-01
	CD8 T cells	4.58E-01	7.19E-01	7.73E-01	0.00E+00	4.53E-01	8.36E-01
	Cycling neutrophil progenitors	1.63E-01	7.69E-01	1.44E-01	0.00E+00	1.60E-01	2.59E-01
	IL1R2 immature neutrophils	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Memory CD4 T cells	1.58E-01	8.65E-01	8.69E-01	0.00E+00	1.54E-01	5.49E-01
	MPO immature neutrophils or progenitors	4.87E-01	9.53E-01	1.69E-01	0.00E+00	4.84E-01	4.82E-01
	Naïve CD8 T cells	3.54E-01	3.84E-01	3.31E-01	0.00E+00	3.49E-01	5.19E-01
	NK cells	6.25E-02	3.62E-01	2.05E-01	0.00E+00	5.97E-02	1.35E-01
Sepsis	**CD8 T cells**	1.61E-04	5.28E-01	6.77E-01	6.02E-01	1.59E-04	5.05E-03
	CD8 T cells	7.52E-08	3.23E-01	4.60E-01	9.98E-01	5.24E-08	2.03E-06
	Cycling neutrophil progenitors	4.56E-05	2.06E-01	8.26E-01	8.03E-01	2.91E-05	4.12E-04
	IL1R2 immature neutrophils	1.68E-10	1.63E-02	7.87E-01	3.74E-01	3.18E-11	5.18E-10
	Memory CD4 T cells	3.29E-03	3.17E-01	9.00E-01	7.31E-01	3.26E-03	2.56E-02
	MPO immature neutrophils or progenitors	8.73E-10	7.60E-01	3.41E-01	9.46E-01	1.29E-09	1.50E-07
	Naïve CD8 T cells	8.64E-09	7.90E-01	9.77E-02	9.36E-01	5.38E-09	2.94E-07
	NK cells	8.54E-01	4.36E-01	2.54E-01	9.81E-01	8.54E-01	6.73E-01
BH Adjusted p values							
Health	**CD8 T cells**	8.92E-01	5.85E-01	6.02E-01	0.00E+00	8.92E-01	6.53E-01
	CD8 T cells	7.34E-01	9.33E-01	9.37E-01	0.00E+00	7.34E-01	9.37E-01
	Cycling neutrophil progenitors	4.35E-01	9.37E-01	4.35E-01	0.00E+00	4.35E-01	5.85E-01
	IL1R2 immature neutrophils	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	Memory CD4 T cells	4.35E-01	9.37E-01	9.37E-01	0.00E+00	4.35E-01	7.86E-01
	MPO immature neutrophils or progenitors	7.35E-01	9.77E-01	4.37E-01	0.00E+00	7.35E-01	7.35E-01
	Naïve CD8 T cells	6.53E-01	6.64E-01	6.53E-01	0.00E+00	6.53E-01	7.68E-01
	NK cells	2.16E-01	6.53E-01	5.03E-01	0.00E+00	2.16E-01	4.31E-01
Sepsis	**CD8 T cells**	8.37E-04	7.68E-01	8.92E-01	8.47E-01	8.37E-04	2.10E-02
	CD8 T cells	6.94E-07	6.53E-01	7.34E-01	9.98E-01	5.43E-07	1.40E-05
	Cycling neutrophil progenitors	2.70E-04	5.03E-01	9.37E-01	9.37E-01	1.86E-04	2.01E-03
	IL1R2 immature neutrophils	6.97E-09	6.45E-02	9.37E-01	6.60E-01	2.64E-09	1.43E-08
	Memory CD4 T cells	1.44E-02	6.53E-01	9.58E-01	9.33E-01	1.44E-02	9.65E-02
	MPO immature neutrophils or progenitors	1.81E-08	9.37E-01	6.53E-01	9.77E-01	2.14E-08	1.25E-06
	Naïve CD8 T cells	1.02E-07	9.37E-01	3.24E-01	9.77E-01	7.44E-08	2.22E-06
	NK cells	9.37E-01	7.34E-01	5.85E-01	9.93E-01	9.37E-01	8.92E-01

Table C. 15. Simple effects tests comparing the effect of time on CIBERSORTx estimated cell fractions in health or sepsis.

Raw p values are reported for each main effect term. Raw p values are also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

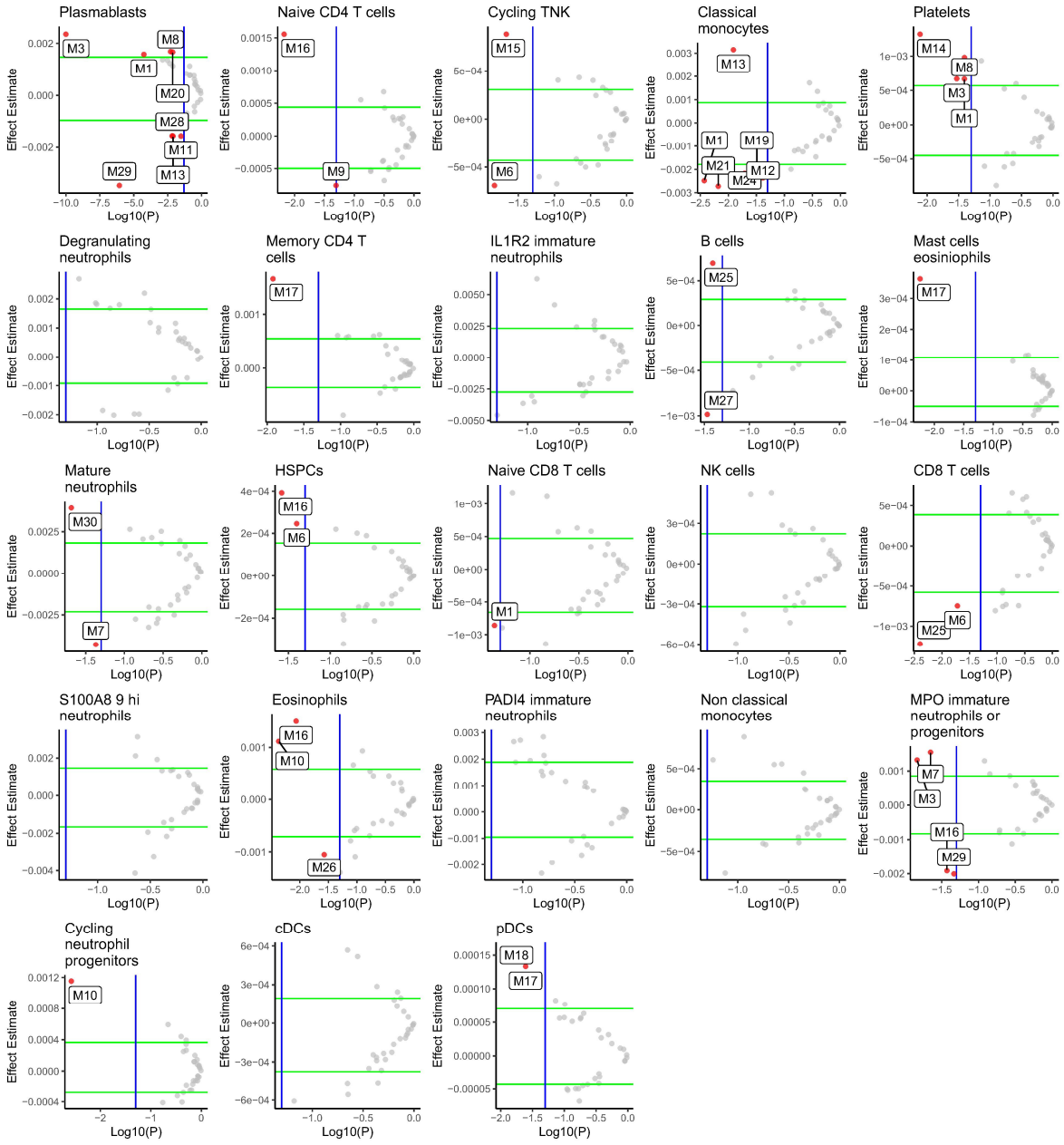


Figure C. 3. Association between BCR module score and CIBERSORTx estimated cell fraction using cell fraction as the dependent variable and module score as the independent variable.

Estimated effect sizes on cell fraction for each module and their corresponding p values are plotted. Log10(p=0.05) is shown in blue and the mean ± 1 SD is shown in green. Modules which meet these thresholds are annotated.

Isotype	Linear Mixed Effect Model						ANOVA	
	DISEASE.SEPSIS	TIMEPOINT	TIMEPOINTx SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values								
IGHA1	6.52E-01	6.22E-01	2.33E-05	3.49E-01	8.83E-01	4.16E-01	4.41E-15	2.87E-13
IGHA2	8.99E-01	6.79E-01	1.40E-04	5.18E-01	8.78E-01	2.67E-01	9.59E-13	5.85E-11
IGHD	1.28E-03	8.60E-01	9.93E-03	1.30E-01	3.63E-01	4.82E-01	7.90E-08	3.68E-08
IGHE	6.58E-02	3.19E-01	6.84E-02	4.07E-01	2.00E-01	4.48E-01	8.66E-07	8.62E-07
IGHG1	3.72E-01	5.61E-01	1.67E-04	3.78E-01	9.62E-01	6.21E-02	3.97E-14	2.68E-12
IGHG2	4.80E-01	5.74E-01	6.75E-04	2.75E-01	6.63E-01	8.22E-02	1.69E-11	2.01E-10
IGHG3	4.96E-01	6.85E-01	5.84E-05	9.36E-01	8.17E-01	5.07E-01	1.63E-13	1.04E-11
IGHG4	3.89E-01	5.26E-01	1.86E-03	4.56E-02	4.57E-01	3.63E-02	1.87E-10	2.77E-10
IGHM	2.06E-03	3.13E-01	2.11E-03	9.98E-01	7.97E-01	1.86E-01	4.00E-13	4.64E-15
BH Adjusted p values								
IGHA1	7.82E-01	7.60E-01	8.83E-05	5.98E-01	9.35E-01	6.23E-01	1.67E-13	4.14E-12
IGHA2	9.39E-01	7.82E-01	4.81E-04	6.76E-01	9.35E-01	5.05E-01	9.86E-12	3.83E-10
IGHD	3.84E-03	9.35E-01	2.55E-02	2.68E-01	6.04E-01	6.67E-01	3.55E-07	1.76E-07
IGHE	1.48E-01	5.59E-01	1.49E-01	6.23E-01	3.90E-01	6.58E-01	3.46E-06	3.46E-06
IGHG1	6.04E-01	7.08E-01	5.47E-04	6.04E-01	9.75E-01	1.44E-01	9.53E-13	2.41E-11
IGHG2	6.67E-01	7.13E-01	2.11E-03	5.08E-01	7.82E-01	1.74E-01	1.21E-10	1.11E-09
IGHG3	6.74E-01	7.82E-01	2.10E-04	9.62E-01	9.06E-01	6.76E-01	2.94E-12	8.31E-11
IGHG4	6.08E-01	6.76E-01	5.35E-03	1.09E-01	6.59E-01	9.01E-02	1.11E-09	1.42E-09
IGHM	5.63E-03	5.59E-01	5.63E-03	9.98E-01	8.96E-01	3.71E-01	4.80E-12	1.67E-13

Table C. 16. Linear mixed effects models comparing the effect of disease status and time point on immunoglobulin isotype usage (bulk RNA sequencing).

P values are reported for each main effect term. P values are also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

Isotype	Day	Simple Effect Tests comparing health and sepsis at each time point				ANOVA
		DISEASE.SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL
Raw p values						
IGHA1	Day1	8.67E-01	2.45E-01	9.18E-01	4.09E-01	8.67E-01
	Day3	4.52E-01	5.04E-01	6.20E-01	3.84E-01	4.52E-01
	Day5	2.10E-01	4.47E-01	9.32E-01	5.04E-01	2.10E-01
IGHA2	Day1	5.16E-01	3.46E-01	7.40E-01	1.87E-01	5.16E-01
	Day3	3.23E-01	7.29E-01	2.76E-01	2.06E-01	3.23E-01
	Day5	2.75E-01	2.83E-01	6.12E-01	4.01E-01	2.75E-01
IGHD	Day1	4.51E-03	2.29E-01	4.72E-01	5.38E-01	4.51E-03
	Day3	1.82E-02	6.57E-02	4.08E-01	3.33E-01	1.82E-02
	Day5	1.83E-01	3.28E-02	3.29E-01	5.25E-01	1.83E-01
IGHG1	Day1	1.98E-01	2.93E-01	9.33E-01	8.49E-02	1.98E-01
	Day3	9.03E-02	3.36E-01	7.24E-01	1.07E-01	9.03E-02
	Day5	2.50E-02	6.59E-01	9.12E-01	5.66E-01	2.50E-02
IGHG2	Day1	9.45E-01	8.48E-02	7.54E-01	7.28E-02	9.45E-01
	Day3	6.08E-01	2.28E-01	5.56E-01	9.90E-02	6.08E-01
	Day5	2.41E-01	1.92E-01	8.91E-01	9.85E-01	2.41E-01
IGHG3	Day1	7.18E-01	6.35E-01	7.34E-01	6.24E-01	7.18E-01
	Day3	5.47E-01	6.90E-01	6.74E-01	2.88E-01	5.47E-01
	Day5	1.15E-01	1.20E-01	7.07E-01	7.01E-01	1.15E-01
IGHG4	Day1	7.07E-01	4.17E-02	5.53E-01	4.03E-02	7.07E-01
	Day3	6.01E-01	7.71E-02	4.72E-01	2.80E-02	6.01E-01
	Day5	8.43E-01	6.41E-01	9.68E-01	8.36E-01	8.43E-01
IGHM	Day1	9.14E-03	9.04E-01	7.28E-01	3.16E-01	9.14E-03
	Day3	2.81E-01	9.96E-01	7.54E-01	1.24E-01	2.81E-01
	Day5	1.54E-01	3.74E-01	7.46E-01	7.65E-01	1.54E-01
BH Adjusted p values						
IGHA1	Day1	9.54E-01	7.16E-01	9.69E-01	8.04E-01	9.54E-01
	Day3	8.47E-01	8.78E-01	8.79E-01	7.94E-01	8.47E-01
	Day5	7.01E-01	8.47E-01	9.69E-01	8.78E-01	7.01E-01
IGHA2	Day1	8.78E-01	7.41E-01	8.79E-01	7.01E-01	8.78E-01
	Day3	7.34E-01	8.79E-01	7.17E-01	7.01E-01	7.34E-01
	Day5	7.17E-01	7.17E-01	8.79E-01	8.04E-01	7.17E-01
IGHD	Day1	2.70E-01	7.16E-01	8.59E-01	8.78E-01	2.70E-01
	Day3	3.64E-01	5.70E-01	8.04E-01	7.34E-01	3.64E-01
	Day5	7.01E-01	3.93E-01	7.34E-01	8.78E-01	7.01E-01
IGHG1	Day1	7.01E-01	7.17E-01	9.69E-01	5.70E-01	7.01E-01
	Day3	5.70E-01	7.34E-01	8.79E-01	5.93E-01	5.70E-01
	Day5	3.74E-01	8.79E-01	9.69E-01	8.79E-01	3.74E-01
IGHG2	Day1	9.69E-01	5.70E-01	8.79E-01	5.70E-01	9.69E-01
	Day3	8.79E-01	7.16E-01	8.78E-01	5.93E-01	8.79E-01
	Day5	7.16E-01	7.01E-01	9.69E-01	9.93E-01	7.16E-01
IGHG3	Day1	8.79E-01	8.79E-01	8.79E-01	8.79E-01	8.79E-01
	Day3	8.78E-01	8.79E-01	8.79E-01	7.17E-01	8.78E-01
	Day5	5.93E-01	5.93E-01	8.79E-01	8.79E-01	5.93E-01
IGHG4	Day1	8.79E-01	4.17E-01	8.78E-01	4.17E-01	8.79E-01
	Day3	8.79E-01	5.70E-01	8.59E-01	3.74E-01	8.79E-01
	Day5	9.45E-01	8.79E-01	9.84E-01	9.45E-01	9.45E-01
IGHM	Day1	2.74E-01	9.69E-01	8.79E-01	7.34E-01	2.74E-01
	Day3	7.17E-01	9.96E-01	8.79E-01	5.93E-01	7.17E-01
	Day5	6.86E-01	7.88E-01	8.79E-01	8.83E-01	6.86E-01

Table C. 17. Simple effects tests comparing the log2(CPM+1) immunoglobulin isotype gene expression between health and sepsis at each time point.

P values are reported for each main effect term. P values are also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

Isotype	Group	Simple Effects Tests comparing the effect of time in either health or sepsis				ANOVA	
		TIME POINT	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p value							
IGHA1	Health	2.47E-01	3.20E-01	3.41E-01	0.00E+00	2.42E-01	4.06E-01
	Sepsis	3.15E-14	2.49E-01	7.37E-01	3.41E-01	8.35E-14	9.99E-12
IGHA2	Health	3.50E-01	3.97E-01	1.94E-01	0.00E+00	3.44E-01	4.06E-01
	Sepsis	3.21E-12	4.38E-01	6.87E-01	2.51E-01	9.38E-12	8.20E-10
IGHD	Health	6.32E-01	7.15E-01	3.08E-01	0.00E+00	6.28E-01	6.31E-01
	Sepsis	1.64E-06	1.51E-01	2.92E-01	4.69E-01	1.75E-06	3.77E-05
IGHG1	Health	2.14E-02	2.43E-02	2.43E-01	0.00E+00	2.00E-02	1.23E-02
	Sepsis	2.23E-12	2.40E-01	8.93E-01	5.38E-02	3.49E-12	1.16E-10
IGHG2	Health	5.56E-02	2.79E-01	6.60E-01	0.00E+00	5.29E-02	1.72E-01
	Sepsis	4.53E-11	1.53E-01	5.94E-01	5.73E-02	9.16E-11	2.16E-09
IGHG3	Health	2.29E-01	1.83E-01	9.10E-01	0.00E+00	2.24E-01	3.17E-01
	Sepsis	6.95E-13	8.09E-01	7.82E-01	4.86E-01	1.57E-12	2.63E-10
IGHG4	Health	1.31E-01	6.07E-01	8.98E-01	0.00E+00	1.27E-01	4.43E-01
	Sepsis	3.90E-10	3.00E-02	5.08E-01	2.92E-02	3.92E-10	3.25E-09
IGHM	Health	6.85E-02	4.71E-01	7.31E-01	0.00E+00	6.54E-02	2.66E-01
	Sepsis	1.42E-11	8.83E-01	8.32E-01	1.98E-01	1.08E-11	7.49E-10
BH Adjusted p value							
IGHA1	Health	4.25E-01	4.85E-01	4.96E-01	0.00E+00	4.25E-01	5.50E-01
	Sepsis	2.77E-12	4.25E-01	8.00E-01	4.96E-01	3.67E-12	9.51E-11
IGHA2	Health	4.96E-01	5.50E-01	3.96E-01	0.00E+00	4.96E-01	5.50E-01
	Sepsis	4.39E-11	5.82E-01	7.75E-01	4.25E-01	9.51E-11	3.80E-09
IGHD	Health	7.32E-01	7.97E-01	4.84E-01	0.00E+00	7.32E-01	7.32E-01
	Sepsis	6.55E-06	3.36E-01	4.67E-01	6.01E-01	6.71E-06	1.38E-04
IGHG1	Health	6.96E-02	7.63E-02	4.25E-01	0.00E+00	6.76E-02	4.32E-02
	Sepsis	3.92E-11	4.25E-01	9.09E-01	1.48E-01	4.39E-11	7.27E-10
IGHG2	Health	1.48E-01	4.54E-01	7.55E-01	0.00E+00	1.48E-01	3.70E-01
	Sepsis	3.32E-10	3.36E-01	7.26E-01	1.48E-01	6.20E-10	9.50E-09
IGHG3	Health	4.25E-01	3.84E-01	9.10E-01	0.00E+00	4.25E-01	4.85E-01
	Sepsis	2.04E-11	8.58E-01	8.39E-01	6.12E-01	3.46E-11	1.54E-09
IGHG4	Health	3.03E-01	7.32E-01	9.09E-01	0.00E+00	3.01E-01	5.82E-01
	Sepsis	2.03E-09	8.80E-02	6.30E-01	8.80E-02	2.03E-09	1.36E-08
IGHM	Health	1.67E-01	6.01E-01	8.00E-01	0.00E+00	1.65E-01	4.42E-01
	Sepsis	1.13E-10	9.09E-01	8.72E-01	3.96E-01	9.51E-11	3.66E-09

Table C. 18. Simple effects tests comparing the effect of time in health or sepsis on the log2(CPM+1) immunoglobulin isotype gene expression: raw p values.

P values are reported for each main effect term. P values are also reported for the results of the ANOVA test, comparing the variance explained by the complete model to the variance explained by the covariate only model. Significant p values ($p<0.05$) are recorded in red. Results are reported to 3 significant figures.

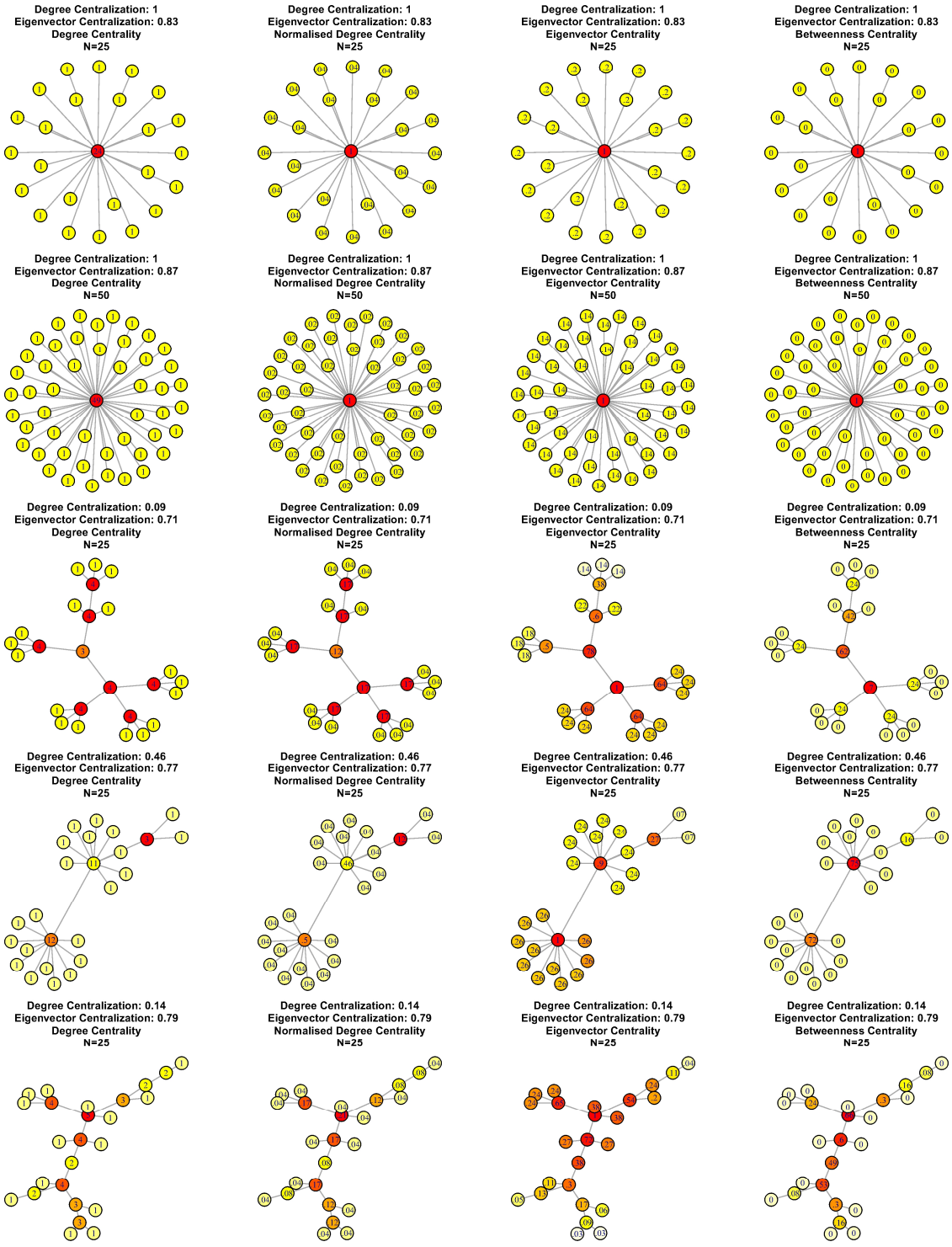


Figure C. 4. Example networks coloured by node level centrality measures from left to right: degree centrality, normalised degree centrality, eigenvector centrality and betweenness centrality.

Graph level centralisation measures (degree and eigenvector) are included in the graph title. For graphs containing the same number of vertexes, network organisation yields diverse centralisation measures.

D Appendices: The TCR Repertoire in Sepsis

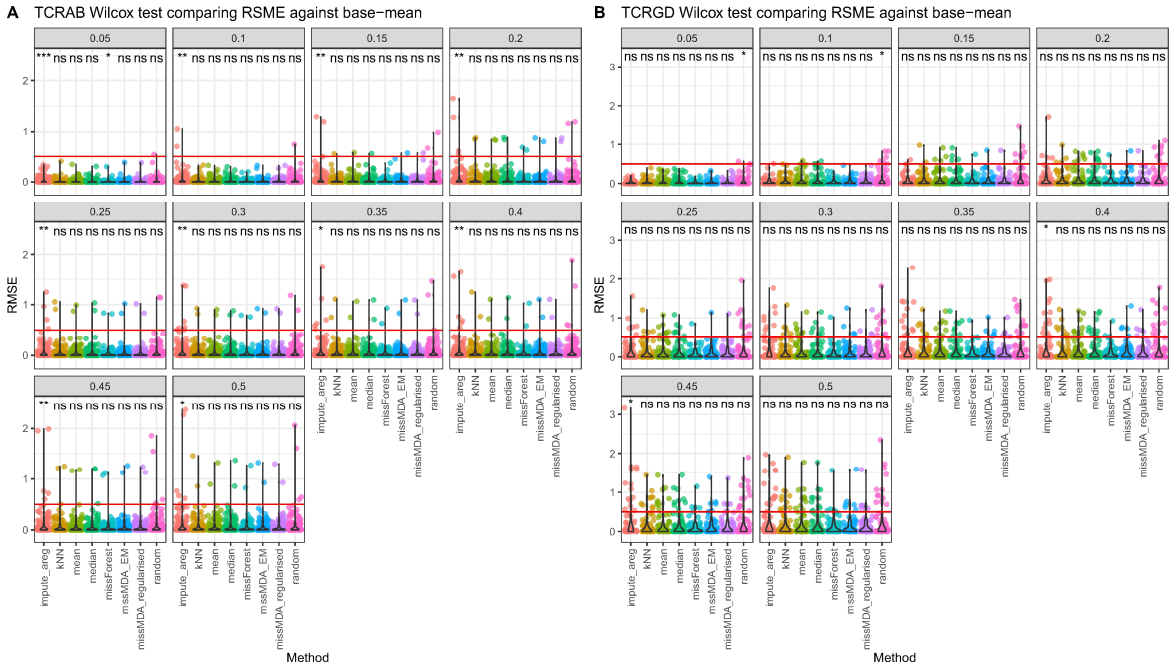


Figure D. 1. Imputation optimisation: missing data imputation methods were evaluated by comparing the root mean squared error (RMSE) for all repertoire metrics at a given NA seed frequency for each TCR receptor repertoire A) TCRαβ and B) TCRδγ.

Results shown are for column-wise (feature) 'NA' seeding. Significance was calculated using a Wilcoxon-test comparing each method to the base-mean of all other groups for each NA seed frequency. The red line indicates RMSE=0.5.

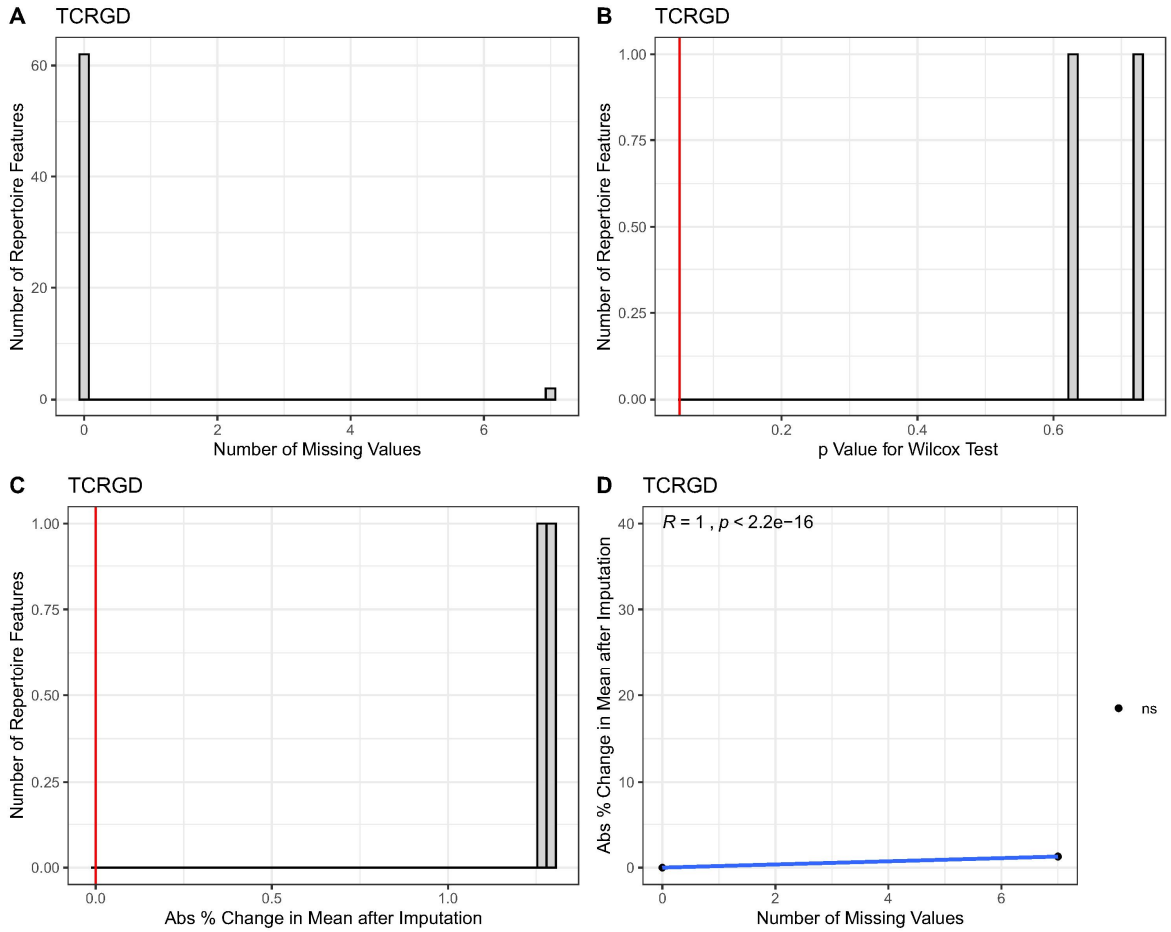


Figure D. 2. TCRGD repertoire imputation summary.

A) The number of missing values for each of 2 TCRGD repertoire features that passed basic filtering. **B)** A Wilcoxon test was used to test for a significant difference in mean after imputation with missForest ($p<0.05$). **C)** The absolute percentage change in mean post-imputation for features with missing data. **D)** There was a significant positive correlation between the number of missing values and the absolute percentage change in mean ($R=0.57$, $p<0.05$).

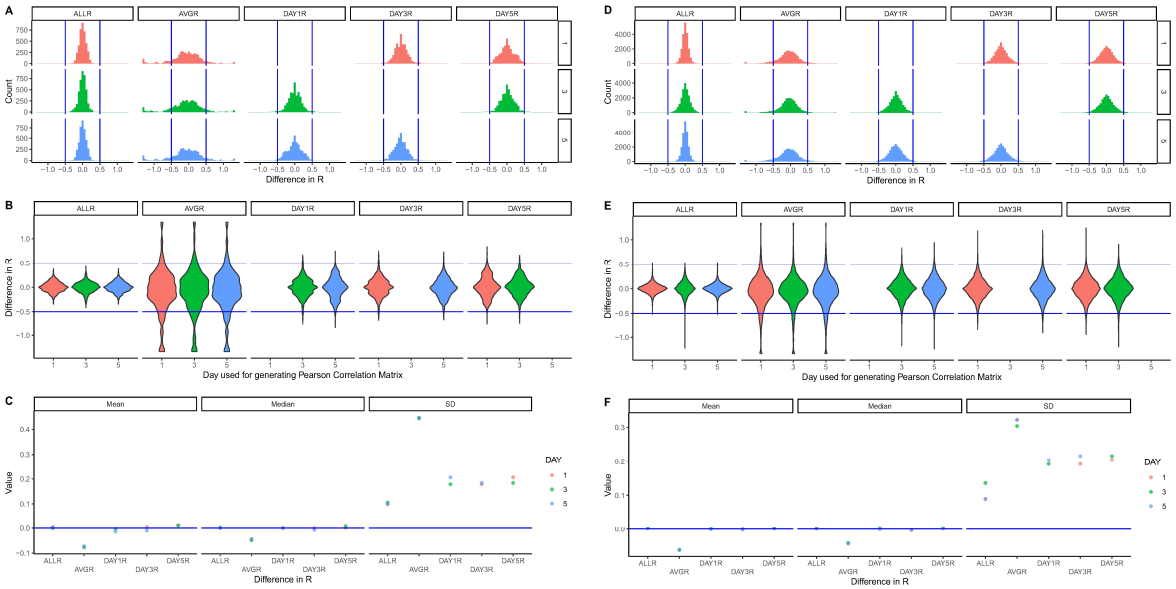


Figure D. 3. Comparing co-expression networks derived using PCCs for day 1, day 3, day 5, all available data (ALLR) and an average of day 1 and day 5 PCCs (AVGR).

Plots **A-C** are for TCRGD, plots **D-F** are for TCRAB. The change in PCC (ΔR) for every pair of features was calculated across methods**. A&D)** Histograms showing the distribution of ΔR between methods. A good method would produce a ΔR distribution centred on 0 (no difference in correlation) with minimal spread. Where the same method is compared ΔR=0 for all feature pairs and has not been visualised. **B&E)** Violin plots conveying the same information as in A. **C&F)** The mean, median and SD for ΔR between different methods.

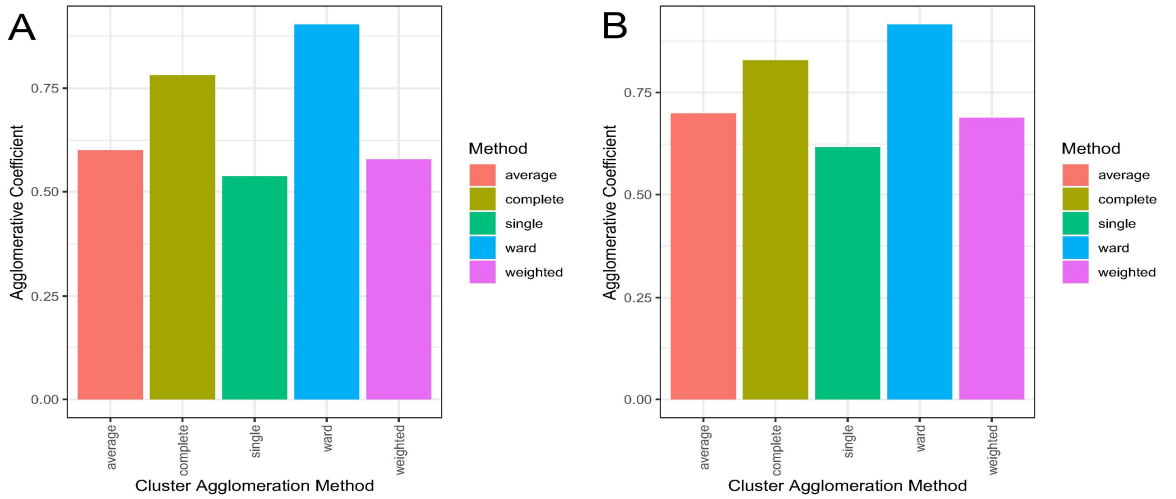


Figure D. 4. Optimal linkage method for A) TCRAB and B) TRGD.

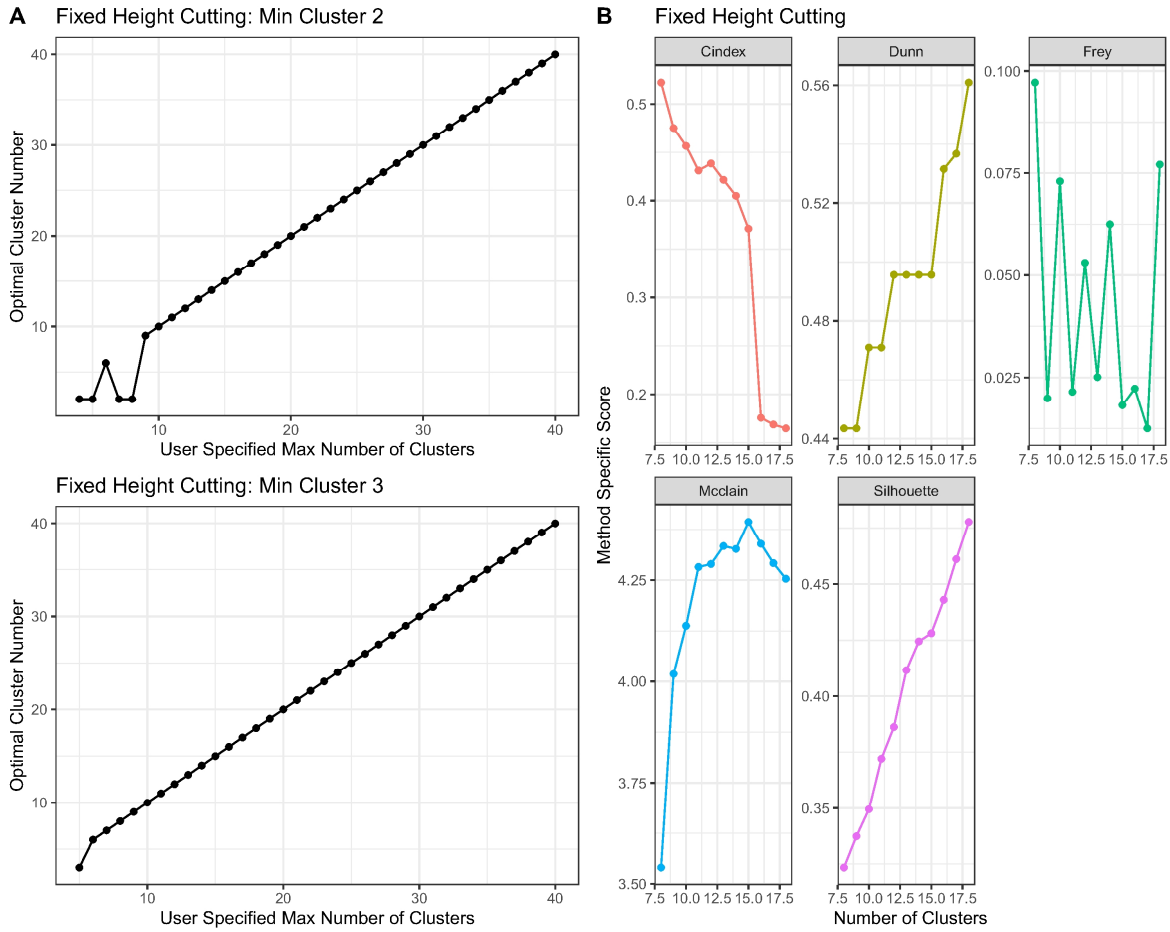


Figure D. 5. TCRGD fixed height cutting.

A,B) The optimal number of clusters proposed by the five cluster detection methods available for hierarchical clustering in 'NBclust' for **A)** a minimum number of clusters of 2 or 3 and variable maximum and **B)** a minimum number of clusters of 3 and variable maximum.

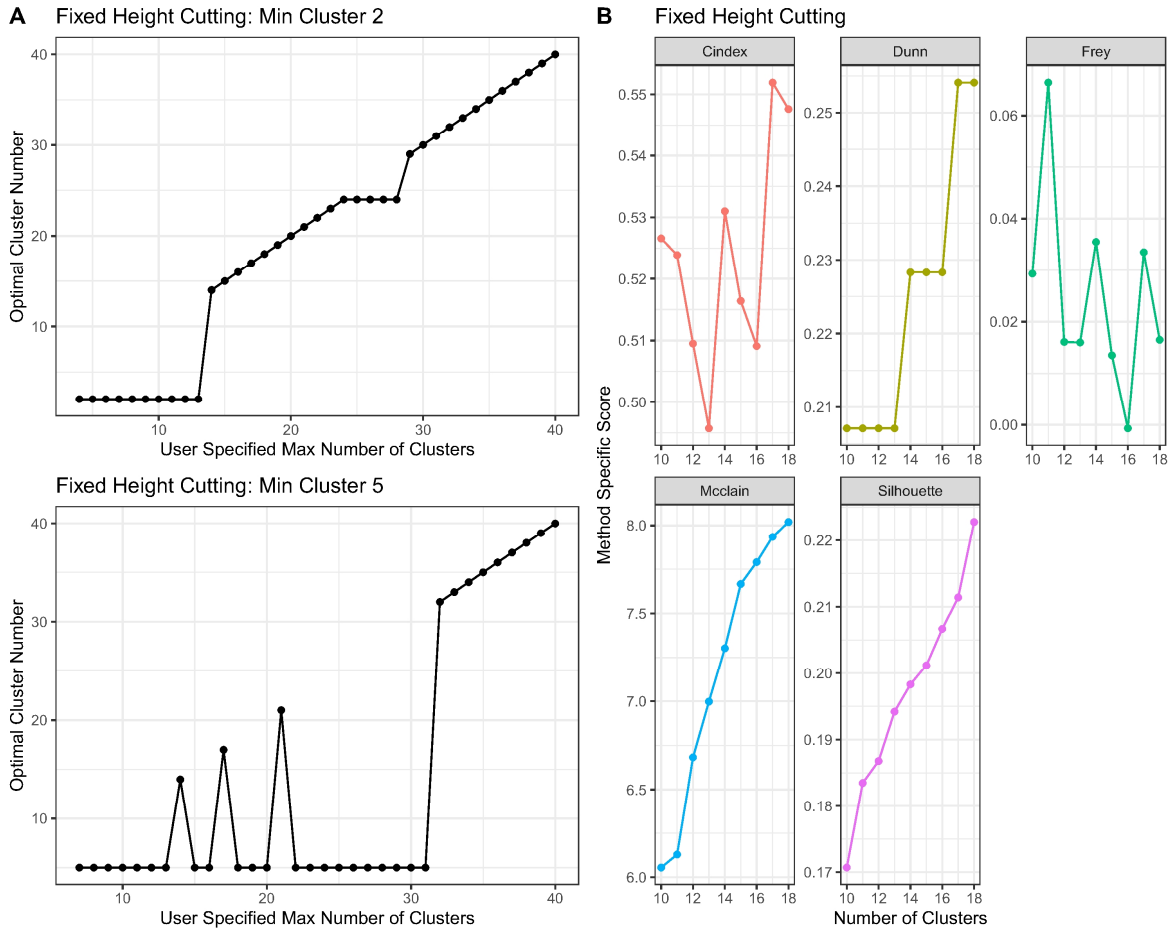


Figure D. 6. TCRAB fixed height cutting.

A, B) The optimal number of clusters proposed by the five cluster detection methods available for hierarchical clustering in 'NBclust' for **A)** a minimum number of clusters of 2 or 5 and variable maximum and **B)** a minimum number of clusters of 5 and variable maximum.

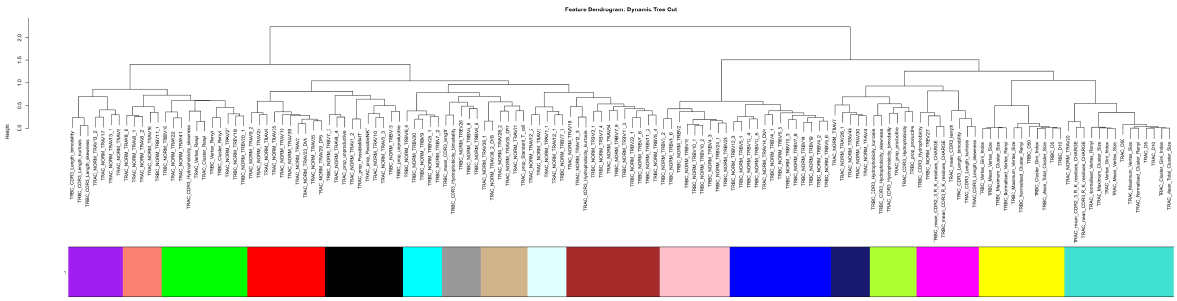


Figure D. 7. TCRAB feature cluster partition using Dynamic Hybrid Tree Cut.

Features in light grey were excluded based on internal filtering parameters.

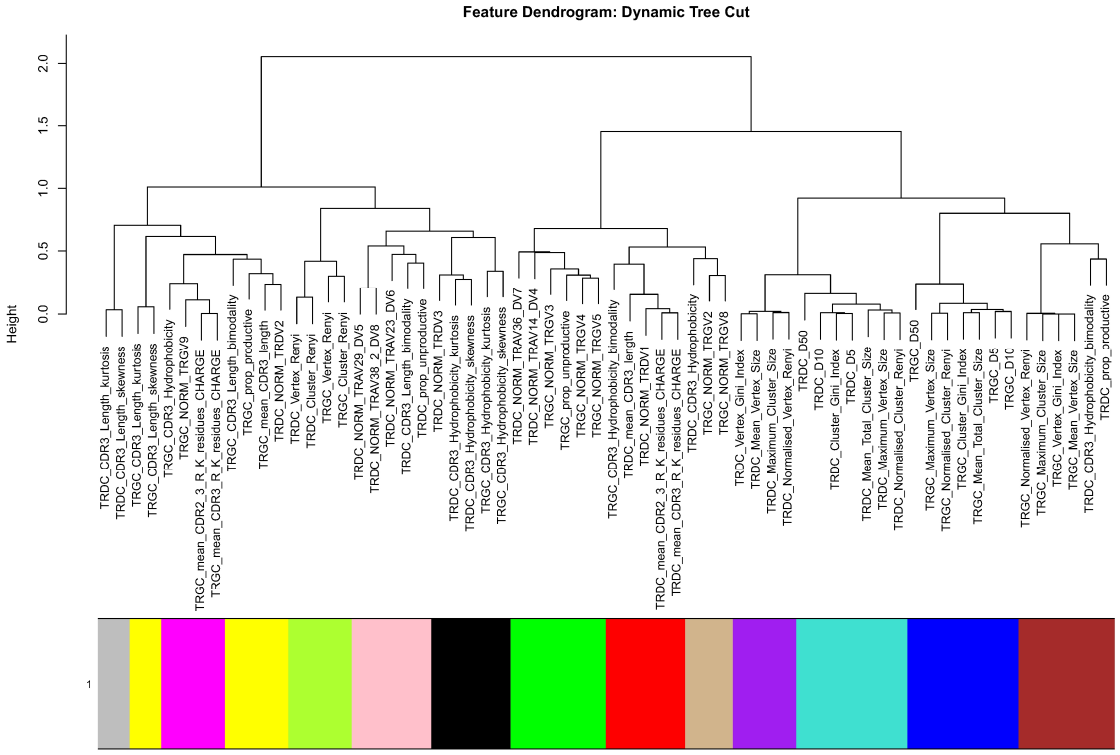


Figure D. 8. TCRGD feature cluster partition using Dynamic Hybrid Tree Cut.

Features in light grey were excluded based on internal filtering parameters.

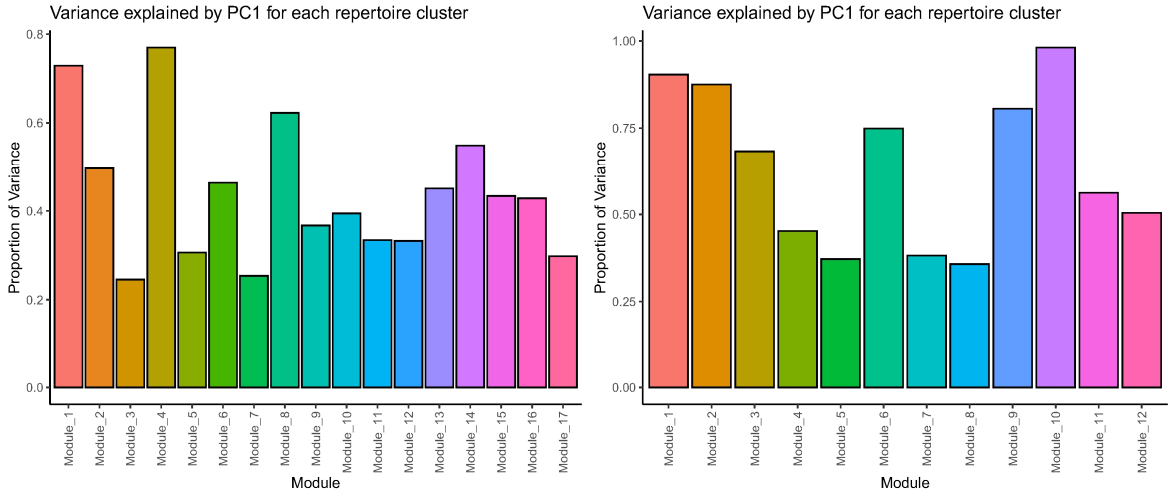


Figure D. 9. Principle component analysis: proportion of variance explained by PC1 for A) TCRAB and B) TCRGD.

For the majority of modules PC1 captures a large proportion of variance and the new eigenvector coordinates well represent trends seen in individual cluster features.

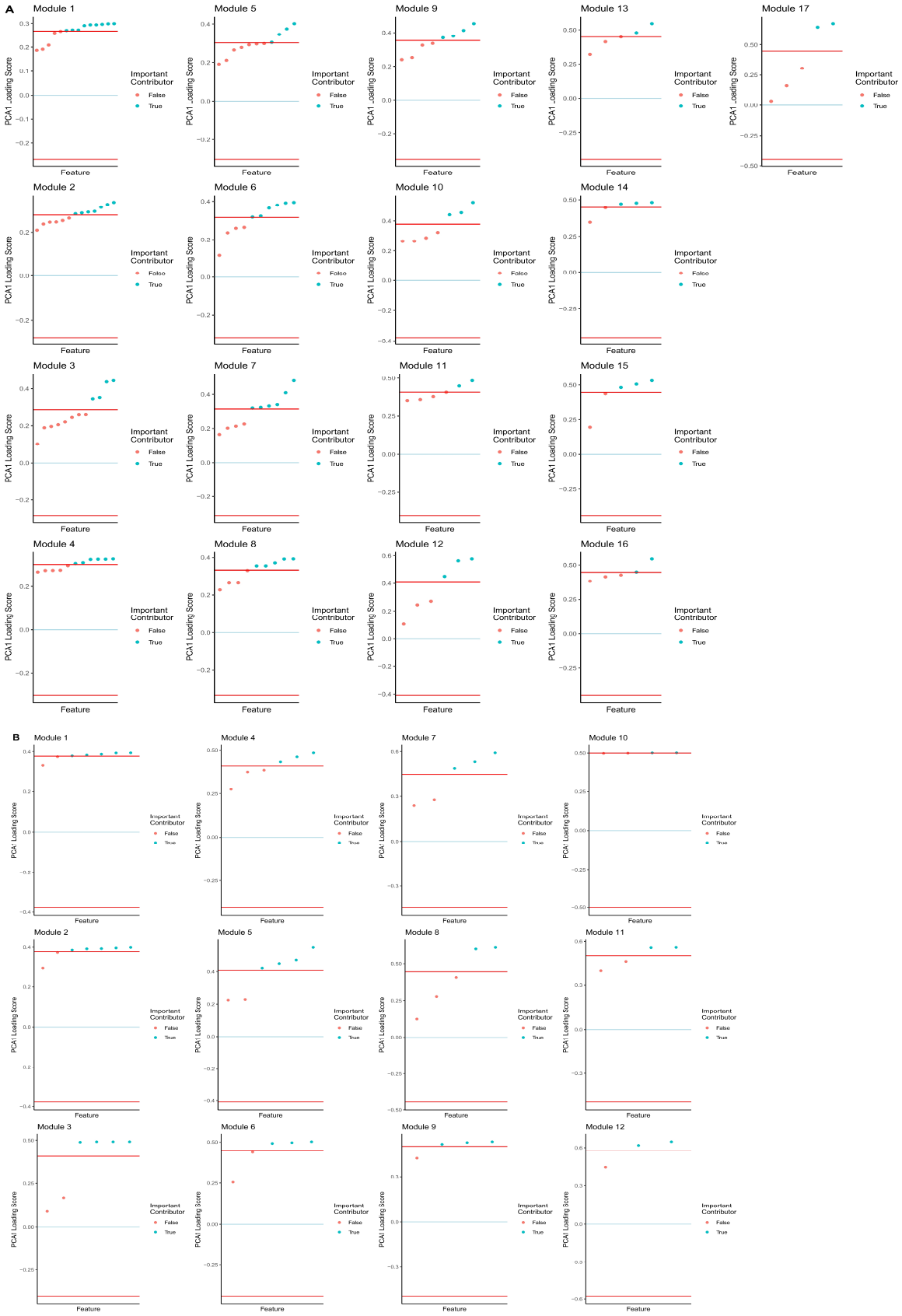


Figure D. 10. PC1 loading scores for A) TCRAB and B) TCRGD.

The module specific 'large loading score' threshold is shown as the red horizontal line and features are classified according to this threshold.

Module	Chain	Model	AICc	AICcWt
Module_1	TCRAB	C	823.425	0.452
Module_2	TCRAB	D*TA	695.039	0.535
Module_3	TCRAB	DTA	640.892	0.244
Module_4	TCRAB	C	794.225	0.205
Module_5	TCRAB	DT	579.584	0.257
Module_6	TCRAB	Null	691.698	0.193
Module_7	TCRAB	DT	483.309	0.194
Module_8	TCRAB	D*TA	410.567	0.291
Module_9	TCRAB	Null	648.255	0.268
Module_10	TCRAB	ASC	661.996	0.269
Module_11	TCRAB	A	538.158	0.191
Module_12	TCRAB	S	558.864	0.327
Module_13	TCRAB	DT	604.032	0.181
Module_14	TCRAB	ASC	637.473	0.657
Module_15	TCRAB	DTS	528.738	0.256
Module_16	TCRAB	DTS	596.791	0.191
Module_17	TCRAB	S	436.938	0.229
Module_1	TCRGD	DTS	644.599	0.158
Module_2	TCRGD	D*TA	643.712	0.284
Module_3	TCRGD	Null	677.934	0.215
Module_4	TCRGD	DT	566.577	0.256
Module_5	TCRGD	DT	469.477	0.218
Module_6	TCRGD	DT	551.686	0.238
Module_7	TCRGD	D*T	475.179	0.292
Module_8	TCRGD	Null	515.098	0.270
Module_9	TCRGD	DT	514.642	0.209
Module_10	TCRGD	C	645.595	0.332
Module_11	TCRGD	C	528.909	0.245
Module_12	TCRGD	C	436.687	0.446

Table D. 1. TCR repertoire: AIC based model selection.

AICc and weighted AICcWt scores are reported for the optimal statistical model for each TCR repertoire module. Model terms are recorded in **Table 23**. Scores are reported to 3 decimal places.

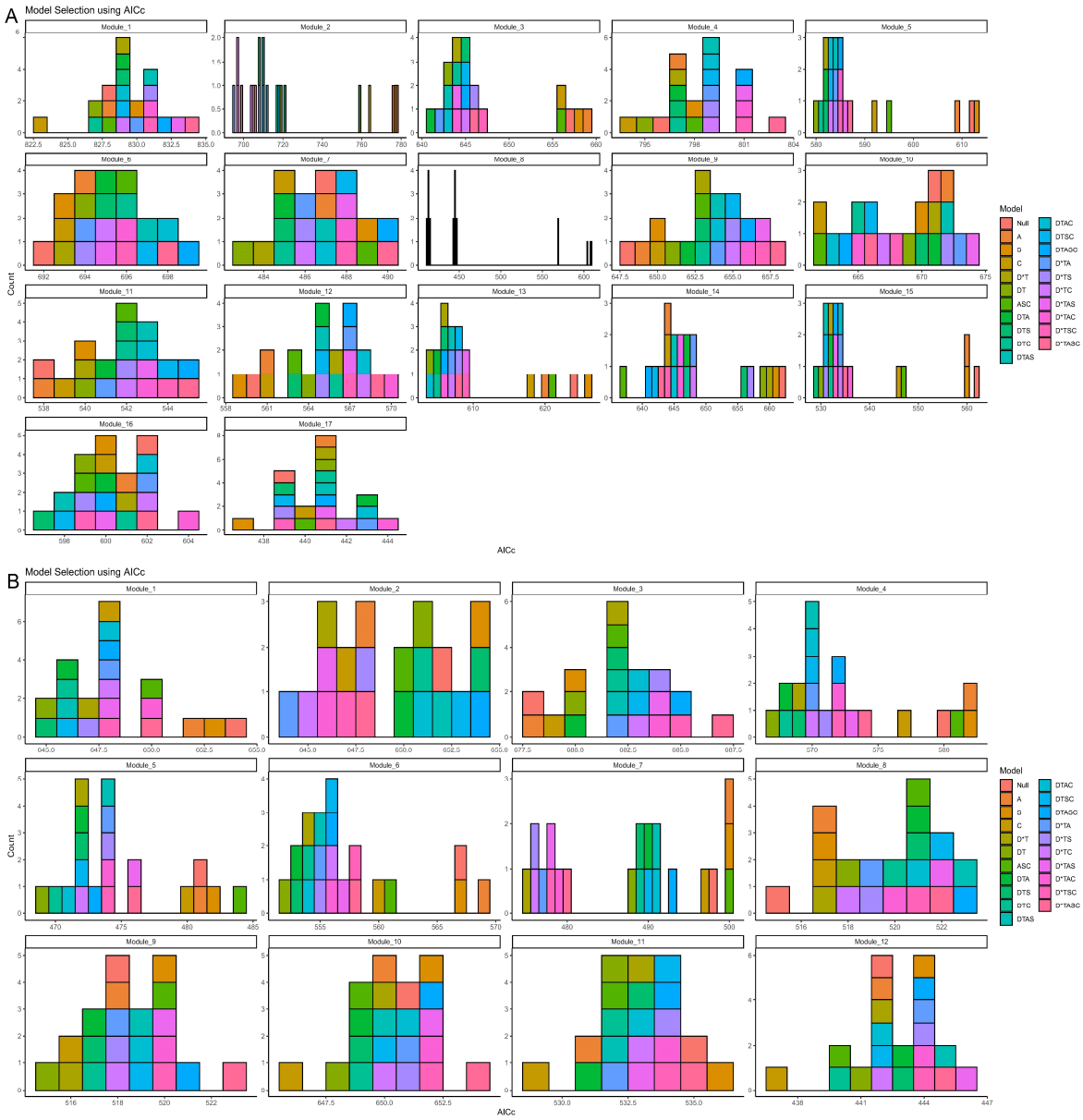


Figure D. 11. AIC model selection for A) TCRAB and B) TCRGD modules.

For each module, the AICc score was calculated for all candidate models, but for most modules AIC model selection did not identify an optimal model.

Module	Receptor	Linear Mixed Effect Model Term						ANOVA	
		DISEASE SEPSIS	TIMEPOINT	TIMEPOINTx SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values									
1	AB	4.72E-01	8.29E-01	6.64E-01	9.61E-01	9.51E-01	2.58E-01	8.83E-01	2.93E-01
2	AB	1.22E-03	2.94E-05	1.05E-04	3.61E-03	5.90E-01	8.99E-01	3.04E-14	2.72E-17
3	AB	1.89E-02	5.06E-02	2.89E-01	1.11E-01	6.82E-01	8.34E-01	9.60E-04	6.00E-04
4	AB	3.16E-01	6.43E-01	6.29E-01	9.47E-01	6.29E-01	7.80E-01	5.52E-01	4.31E-01
5	AB	1.61E-03	7.94E-01	6.29E-01	5.50E-01	8.14E-01	2.52E-01	2.39E-03	1.29E-06
6	AB	5.01E-02	1.03E-01	6.30E-02	4.04E-01	3.63E-01	9.22E-01	1.83E-01	3.21E-01
7	AB	1.60E-01	1.98E-01	2.39E-01	6.10E-01	8.39E-01	8.38E-01	1.58E-01	1.57E-01
8	AB	6.34E-22	3.91E-06	7.94E-09	1.32E-01	8.79E-01	6.75E-01	1.08E-34	8.26E-42
9	AB	5.25E-01	6.53E-01	7.70E-01	2.42E-01	6.89E-01	5.12E-01	7.84E-01	8.18E-01
10	AB	5.07E-01	5.63E-01	9.43E-01	4.62E-02	3.70E-01	1.87E-03	5.49E-01	9.31E-03
11	AB	8.46E-02	1.72E-01	1.49E-01	2.49E-01	8.93E-01	5.16E-01	3.01E-01	4.32E-01
12	AB	7.08E-01	7.78E-01	6.63E-01	8.77E-01	8.65E-02	6.37E-01	9.66E-01	7.23E-01
13	AB	3.54E-03	5.09E-01	5.81E-01	7.02E-02	9.78E-01	7.56E-02	2.08E-04	6.64E-05
14	AB	7.60E-01	2.70E-01	3.39E-01	6.87E-02	1.49E-02	2.04E-02	6.92E-01	1.49E-05
15	AB	7.93E-04	6.39E-01	5.37E-01	8.06E-01	6.56E-02	9.75E-01	3.34E-04	5.01E-07
16	AB	3.56E-01	7.79E-01	5.05E-01	6.99E-01	3.24E-02	4.90E-01	2.56E-01	3.80E-02
17	AB	1.46E-01	6.04E-01	1.43E-01	4.67E-01	3.05E-02	1.32E-01	1.35E-01	1.06E-01
1	GD	8.10E-02	3.39E-01	9.99E-01	4.27E-01	1.46E-01	8.73E-01	7.99E-02	8.51E-03
2	GD	1.52E-01	4.49E-02	5.13E-03	1.84E-01	8.52E-01	5.71E-01	3.05E-02	8.66E-03
3	GD	6.65E-01	2.14E-01	5.31E-01	3.13E-01	7.41E-01	9.24E-01	4.89E-01	6.17E-01
4	GD	4.13E-03	4.44E-01	9.11E-01	8.16E-01	9.79E-01	5.23E-01	3.97E-03	4.82E-03
5	GD	2.93E-04	9.30E-01	8.80E-01	3.54E-01	6.14E-01	1.06E-01	2.13E-03	5.45E-03
6	GD	2.88E-02	5.30E-01	6.84E-01	1.86E-01	6.40E-01	3.30E-01	1.96E-02	1.41E-03
7	GD	1.70E-01	1.77E-04	1.19E-04	9.12E-01	2.81E-01	5.36E-01	5.95E-06	1.78E-05
8	GD	9.90E-01	2.12E-01	6.73E-02	4.38E-01	7.60E-01	2.10E-01	1.42E-01	4.25E-01
9	GD	1.42E-01	4.00E-01	4.33E-01	6.95E-01	9.40E-01	9.74E-01	2.72E-01	2.10E-01
10	GD	1.67E-01	4.91E-01	3.62E-01	6.32E-01	3.04E-01	5.07E-01	5.76E-01	9.52E-02
11	GD	1.29E-01	1.81E-01	2.46E-01	2.22E-01	4.00E-01	6.71E-01	3.84E-01	3.77E-02
12	GD	9.54E-01	8.61E-01	4.56E-01	4.93E-01	5.31E-01	7.47E-02	6.96E-01	1.56E-01
BH Adjusted p Values									
1	AB	8.34E-01	9.28E-01	8.85E-01	9.78E-01	9.78E-01	5.66E-01	9.54E-01	6.13E-01
2	AB	8.76E-03	3.63E-04	1.09E-03	2.05E-02	8.81E-01	9.55E-01	8.26E-13	9.24E-16
3	AB	9.54E-02	2.02E-01	6.13E-01	3.44E-01	8.87E-01	9.28E-01	7.25E-03	5.10E-03
4	AB	6.42E-01	8.84E-01	8.84E-01	9.78E-01	8.84E-01	9.28E-01	8.53E-01	7.83E-01
5	AB	1.09E-02	9.28E-01	8.84E-01	8.53E-01	9.28E-01	5.66E-01	1.48E-02	2.19E-05
6	AB	2.02E-01	3.34E-01	2.45E-01	7.53E-01	6.96E-01	9.72E-01	4.52E-01	6.42E-01
7	AB	4.11E-01	4.81E-01	5.66E-01	8.84E-01	9.28E-01	9.28E-01	4.11E-01	4.11E-01
8	AB	2.87E-20	5.91E-05	1.80E-07	3.90E-01	9.54E-01	8.87E-01	7.32E-33	1.12E-39
9	AB	8.50E-01	8.85E-01	9.28E-01	5.66E-01	8.87E-01	8.45E-01	9.28E-01	9.28E-01
10	AB	8.45E-01	8.61E-01	9.78E-01	1.96E-01	6.98E-01	1.21E-02	8.53E-01	5.06E-02
11	AB	2.87E-01	4.33E-01	4.06E-01	5.66E-01	9.55E-01	8.45E-01	6.21E-01	7.83E-01
12	AB	8.91E-01	9.28E-01	8.85E-01	9.54E-01	2.87E-01	8.84E-01	9.78E-01	9.02E-01
13	AB	2.05E-02	8.45E-01	8.78E-01	2.51E-01	9.78E-01	2.64E-01	2.02E-03	7.52E-04
14	AB	9.28E-01	5.84E-01	6.68E-01	2.51E-01	7.81E-02	9.92E-02	8.87E-01	2.03E-04
15	AB	6.34E-03	8.84E-01	8.53E-01	9.28E-01	2.48E-01	9.78E-01	3.03E-03	9.73E-06
16	AB	6.91E-01	9.28E-01	8.45E-01	8.88E-01	1.47E-01	8.45E-01	5.66E-01	1.67E-01
17	AB	4.06E-01	8.84E-01	4.04E-01	8.34E-01	1.43E-01	3.90E-01	3.90E-01	3.36E-01
1	GD	3.38E-01	6.77E-01	9.99E-01	7.35E-01	4.82E-01	9.82E-01	3.38E-01	5.94E-02
2	GD	4.82E-01	2.27E-01	4.36E-02	4.97E-01	9.82E-01	7.90E-01	1.72E-01	5.94E-02
3	GD	8.46E-01	5.12E-01	7.57E-01	6.54E-01	8.89E-01	9.91E-01	7.57E-01	8.22E-01
4	GD	4.36E-02	7.35E-01	9.91E-01	9.56E-01	9.99E-01	7.57E-01	4.36E-02	4.36E-02
5	GD	5.63E-03	9.91E-01	9.82E-01	6.94E-01	8.22E-01	4.06E-01	2.92E-02	4.36E-02
6	GD	1.72E-01	7.57E-01	8.46E-01	4.97E-01	8.30E-01	6.73E-01	1.26E-01	2.26E-02
7	GD	4.94E-01	4.26E-03	3.82E-03	9.91E-01	6.12E-01	7.57E-01	5.71E-04	8.56E-04
8	GD	9.99E-01	5.12E-01	3.23E-01	7.35E-01	9.00E-01	5.12E-01	4.82E-01	7.35E-01
9	GD	4.82E-01	7.25E-01	7.35E-01	8.46E-01	9.91E-01	9.99E-01	6.08E-01	5.12E-01
10	GD	4.94E-01	7.57E-01	6.96E-01	8.30E-01	6.49E-01	7.57E-01	7.90E-01	3.81E-01
11	GD	4.77E-01	4.97E-01	5.63E-01	5.19E-01	7.25E-01	8.46E-01	7.22E-01	2.01E-01
12	GD	9.96E-01	9.82E-01	7.42E-01	7.57E-01	7.57E-01	3.38E-01	8.46E-01	4.82E-01

Table D. 2. Linear mixed effects models comparing the effect of time and disease status on TCR module score.

P values are reported for each main effect term in the health versus sepsis linear mixed effects models (D*TASC model). Additionally, p values are provided for the ANOVA test, which compares the variance explained by the D*TASC model to that explained by the null model or the covariate-only model. All p values are reported to 3 significant figures.

Receptor	Module	Time	Simple Effects Tests				ANOVA
			DISEASE SEPSIS	AGE	SEX MALE	CI	COV ONLY MODEL
Raw p values							
AB	2	Day1	1.10E-04	1.17E-02	7.70E-01	5.16E-01	1.10E-04
		Day3	3.05E-05	3.95E-02	9.57E-01	1.20E-01	3.05E-05
		Day5	6.40E-14	2.98E-01	7.43E-01	7.91E-01	6.40E-14
	8	Day1	3.31E-20	1.53E-01	9.34E-01	9.73E-01	3.31E-20
		Day3	3.56E-15	4.45E-01	6.24E-01	1.45E-01	3.56E-15
		Day5	6.59E-21	5.12E-01	8.24E-01	7.52E-01	6.59E-21
GD	2	Day1	1.91E-01	3.48E-01	7.48E-01	7.15E-01	1.91E-01
		Day3	8.01E-01	1.93E-01	8.66E-01	9.46E-01	8.01E-01
		Day5	6.35E-01	2.05E-01	7.23E-01	7.33E-01	6.35E-01
	7	Day1	9.13E-02	9.52E-01	8.30E-01	7.06E-01	9.13E-02
		Day3	3.47E-03	5.20E-01	2.63E-01	7.26E-01	3.47E-03
		Day5	7.07E-05	8.74E-01	3.42E-02	9.88E-01	7.07E-05
BH Adjusted p values							
AB	2	Day1	2.76E-04	2.70E-02	9.13E-01	7.38E-01	2.76E-04
		Day3	9.14E-05	8.47E-02	9.73E-01	2.40E-01	9.14E-05
		Day5	2.40E-13	4.96E-01	9.13E-01	9.13E-01	2.40E-13
	8	Day1	2.48E-19	2.70E-01	9.73E-01	9.73E-01	2.48E-19
		Day3	1.78E-14	7.03E-01	8.50E-01	2.70E-01	1.78E-14
		Day5	9.88E-20	7.38E-01	9.16E-01	9.13E-01	9.88E-20
GD	2	Day1	5.60E-01	8.04E-01	9.71E-01	9.71E-01	5.60E-01
		Day3	9.71E-01	5.60E-01	9.71E-01	9.85E-01	9.71E-01
		Day5	9.71E-01	5.60E-01	9.71E-01	9.71E-01	9.71E-01
	7	Day1	3.91E-01	9.85E-01	9.71E-01	9.71E-01	3.91E-01
		Day3	2.60E-02	9.71E-01	6.57E-01	9.71E-01	2.60E-02
		Day5	1.06E-03	9.71E-01	2.05E-01	9.88E-01	1.06E-03

Table D. 3. Simple Effects tests exploring the effect of disease status (health or sepsis) at each time point on TCR module score.

P values are reported for each main effect term. Additionally, p values are provided for the ANOVA test, which compares the variance explained by the model to that explained by the covariate-only model. All p values are reported to 3 significant figures.

Receptor	Module	Group	Simple Effects Tests				ANOVA	
			TIMEPOINT	AGE	SEX.MALE	COMORBIDITIES.CI	COV ONLY MODEL	NULL MODEL
Raw p values								
AB	2	Health	7.12E-03	3.13E-04	8.02E-02		6.74E-03	4.35E-04
		Sepsis	7.29E-01	1.29E-01	7.60E-02	1.65E-01	7.28E-01	8.24E-03
	8	Health	1.86E-03	4.78E-01	1.19E-01		1.73E-03	8.21E-03
		Sepsis	4.96E-06	1.36E-01	2.47E-01	7.08E-01	3.03E-06	2.68E-05
GD	2	Health	9.87E-03	1.37E-01	5.51E-01		9.14E-03	1.85E-02
		Sepsis	4.87E-02	2.67E-01	9.19E-01	6.18E-01	4.64E-02	1.15E-01
	7	Health	2.27E-02	1.90E-01	2.84E-03		2.14E-02	2.15E-03
		Sepsis	1.28E-01	8.11E-01	6.48E-01	3.17E-01	1.25E-01	4.32E-01
BH Adjusted p values								
AB	2	Health	1.65E-02	1.72E-03	1.36E-01		1.65E-02	1.92E-03
		Sepsis	7.29E-01	1.87E-01	1.36E-01	2.14E-01	7.29E-01	1.65E-02
	8	Health	5.85E-03	5.53E-01	1.87E-01		5.85E-03	1.65E-02
		Sepsis	5.45E-05	1.87E-01	3.02E-01	7.29E-01	5.45E-05	1.96E-04
GD	2	Health	5.43E-02	2.32E-01	6.73E-01	0.00E+00	5.43E-02	7.15E-02
		Sepsis	1.19E-01	3.91E-01	9.19E-01	7.13E-01	1.19E-01	2.32E-01
	7	Health	7.15E-02	2.98E-01	3.12E-02	0.00E+00	7.15E-02	3.12E-02
		Sepsis	2.32E-01	8.49E-01	7.13E-01	4.36E-01	2.32E-01	5.59E-01

Table D. 4. Simple Effects tests exploring the effect of time in health or sepsis on TCR module score.

P values are reported for each main effect term. Additionally, p values are provided for the ANOVA test, which compares the variance explained by the model to that explained by the null model or the covariate-only model. All p values are reported to 3 significant figures.

Receptor	Module	TIMEPOINT	Linear Mixed Effects Model							ANOVA	
			EARLY DEATH	LATE DEATH	TIMEx EARLY DEATH	TIMEx LATE DEATH	AGE	SEX.MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values											
AB	1	8.79E-01	6.73E-01	7.07E-02	6.97E-01	3.46E-01	8.63E-01	8.41E-01	2.28E-01	4.52E-01	6.16E-01
AB	2	5.84E-01	4.83E-02	4.12E-01	2.96E-01	7.75E-01	1.76E-01	8.05E-02	1.98E-01	4.66E-01	1.97E-02
AB	3	1.50E-01	1.49E-01	5.42E-01	1.05E-01	4.63E-01	7.59E-02	5.58E-01	5.88E-01	1.08E-02	1.63E-02
AB	4	6.24E-01	2.02E-01	2.57E-01	3.98E-01	5.72E-01	8.95E-01	5.49E-01	7.48E-01	8.11E-01	9.51E-01
AB	5	9.41E-01	3.10E-01	7.18E-01	2.92E-01	7.15E-01	6.41E-01	7.98E-01	2.86E-01	8.22E-01	9.23E-01
AB	6	4.65E-01	4.15E-01	1.99E-01	5.15E-01	6.75E-01	4.30E-01	3.99E-01	8.44E-01	1.76E-01	3.38E-01
AB	7	4.72E-01	1.37E-04	3.61E-02	6.46E-01	9.61E-03	7.18E-01	9.04E-01	5.06E-01	1.31E-05	1.77E-04
AB	8	8.20E-03	1.44E-03	7.92E-01	1.56E-01	8.76E-01	1.77E-01	3.03E-01	4.60E-01	3.05E-06	7.47E-06
AB	9	9.50E-01	1.73E-02	3.71E-01	6.72E-01	4.87E-01	2.37E-01	4.61E-01	2.23E-01	4.81E-02	1.13E-01
AB	10	3.50E-01	5.66E-01	7.78E-01	7.18E-01	9.35E-01	1.11E-01	1.68E-01	5.72E-03	8.92E-01	1.48E-01
AB	11	3.35E-01	3.50E-01	9.47E-02	1.59E-03	2.91E-01	2.47E-01	8.11E-01	6.57E-01	2.10E-02	6.53E-02
AB	12	5.65E-01	1.14E-01	2.82E-01	6.13E-01	9.22E-01	9.69E-01	1.55E-01	7.50E-01	5.20E-01	5.97E-01
AB	13	4.84E-01	4.51E-01	2.43E-01	4.14E-01	2.78E-01	9.46E-02	8.20E-01	8.48E-02	8.87E-01	7.20E-01
AB	14	7.50E-02	1.12E-04	1.23E-01	6.93E-02	3.57E-01	3.31E-01	9.57E-03	2.50E-02	2.77E-03	6.02E-06
AB	15	7.88E-01	2.00E-01	1.12E-01	9.32E-01	4.21E-01	8.33E-01	1.11E-01	6.97E-01	1.33E-01	1.80E-01
AB	16	9.57E-01	1.45E-05	9.12E-01	4.11E-02	8.65E-01	8.89E-01	3.08E-02	7.81E-01	2.02E-04	1.77E-04
AB	17	8.67E-01	3.36E-03	1.48E-01	9.22E-02	2.70E-01	5.91E-01	1.05E-02	9.04E-02	2.70E-02	1.32E-02
GD	1	1.43E-01	4.11E-02	9.96E-01	7.46E-01	8.64E-01	4.69E-01	2.43E-01	8.62E-01	1.69E-01	2.37E-01
GD	2	3.75E-02	9.28E-02	1.11E-01	4.41E-01	3.51E-01	1.49E-01	8.44E-01	1.00E+00	6.89E-03	1.27E-02
GD	3	8.08E-01	1.30E-01	6.72E-01	5.97E-01	8.97E-01	3.86E-01	9.20E-01	7.69E-01	5.10E-01	7.12E-01
GD	4	8.06E-01	4.72E-01	3.61E-01	1.32E-01	4.25E-01	6.19E-01	9.71E-01	7.60E-01	4.71E-01	7.00E-01
GD	5	8.83E-01	1.02E-02	6.84E-01	6.66E-01	5.07E-01	6.25E-01	5.91E-01	3.92E-01	1.52E-02	4.38E-02
GD	6	1.63E-02	5.67E-02	1.40E-02	2.81E-01	1.67E-01	1.08E-01	9.55E-01	2.81E-01	6.42E-02	1.31E-01
GD	7	5.27E-01	2.50E-04	8.59E-01	1.86E-02	5.38E-01	8.42E-01	5.93E-01	5.05E-01	2.11E-03	9.48E-03
GD	8	8.10E-01	1.10E-02	1.06E-01	6.50E-01	3.94E-01	2.45E-01	6.87E-01	2.89E-01	2.17E-02	5.78E-02
GD	9	5.38E-01	3.51E-01	3.72E-01	5.17E-01	3.88E-01	8.56E-01	9.01E-01	7.06E-01	4.92E-01	7.97E-01
GD	10	3.48E-01	7.89E-01	6.65E-01	8.43E-01	3.71E-01	7.19E-01	3.48E-01	5.46E-01	9.36E-01	8.95E-01
GD	11	8.36E-01	3.79E-01	4.42E-01	9.51E-01	5.41E-01	2.69E-01	4.27E-01	8.80E-01	7.17E-01	6.10E-01
GD	12	3.74E-02	1.74E-03	2.13E-01	1.56E-01	2.32E-01	5.33E-01	2.82E-01	1.57E-01	3.45E-02	5.59E-02
BH Adjusted p values											
AB	1	9.57E-01	8.96E-01	3.44E-01	9.07E-01	6.84E-01	9.57E-01	9.57E-01	5.70E-01	7.69E-01	8.65E-01
AB	2	8.51E-01	2.57E-01	7.50E-01	6.36E-01	9.48E-01	4.94E-01	3.60E-01	5.21E-01	7.69E-01	1.40E-01
AB	3	4.64E-01	4.64E-01	8.37E-01	3.97E-01	7.69E-01	3.49E-01	8.43E-01	8.51E-01	9.19E-02	1.26E-01
AB	4	8.70E-01	5.21E-01	6.07E-01	7.45E-01	8.46E-01	9.57E-01	8.40E-01	9.31E-01	9.50E-01	9.63E-01
AB	5	9.63E-01	6.50E-01	9.07E-01	6.36E-01	9.07E-01	8.85E-01	9.48E-01	6.36E-01	9.50E-01	9.63E-01
AB	6	7.69E-01	7.50E-01	5.21E-01	8.10E-01	8.96E-01	7.62E-01	7.45E-01	9.57E-01	4.94E-01	6.83E-01
AB	7	7.72E-01	3.32E-03	2.12E-01	8.85E-01	9.08E-02	9.07E-01	9.61E-01	8.04E-01	4.94E-04	3.35E-03
AB	8	8.71E-02	2.22E-02	9.48E-01	4.66E-01	9.57E-01	4.94E-01	6.44E-01	7.69E-01	4.23E-04	4.23E-04
AB	9	9.63E-01	1.28E-01	7.08E-01	8.96E-01	7.82E-01	5.84E-01	7.69E-01	5.65E-01	2.57E-01	3.97E-01
AB	10	6.84E-01	8.43E-01	9.48E-01	9.07E-01	9.63E-01	3.97E-01	4.91E-01	6.49E-02	9.57E-01	4.64E-01
AB	11	6.83E-01	6.84E-01	3.74E-01	2.26E-02	6.36E-01	5.90E-01	9.50E-01	8.93E-01	1.43E-01	3.36E-01
AB	12	8.43E-01	3.97E-01	6.36E-01	8.65E-01	9.63E-01	9.69E-01	4.66E-01	9.31E-01	8.10E-01	8.52E-01
AB	13	7.82E-01	7.69E-01	5.90E-01	7.50E-01	6.36E-01	3.74E-01	9.50E-01	3.70E-01	9.57E-01	9.07E-01
AB	14	3.49E-01	3.17E-03	4.19E-01	3.44E-01	6.89E-01	6.83E-01	9.08E-02	1.63E-01	3.62E-02	4.23E-04
AB	15	9.48E-01	5.21E-01	3.97E-01	9.63E-01	7.53E-01	9.57E-01	3.97E-01	9.07E-01	4.45E-01	4.95E-01
AB	16	9.63E-01	4.94E-04	9.63E-01	2.33E-01	9.57E-01	9.57E-01	1.87E-01	9.48E-01	3.43E-03	3.35E-03
AB	17	9.57E-01	4.08E-02	4.64E-01	3.74E-01	6.29E-01	8.51E-01	9.19E-02	3.74E-01	1.70E-01	1.07E-01
GD	1	5.70E-01	2.90E-01	1.00E+00	9.57E-01	9.57E-01	8.74E-01	7.34E-01	9.57E-01	5.80E-01	7.34E-01
GD	2	2.81E-01	4.84E-01	5.11E-01	8.69E-01	8.30E-01	5.70E-01	9.57E-01	1.00E+00	1.78E-01	1.78E-01
GD	3	9.57E-01	5.47E-01	9.38E-01	9.19E-01	9.57E-01	8.30E-01	9.68E-01	9.57E-01	8.74E-01	9.38E-01
GD	4	9.57E-01	8.74E-01	8.30E-01	5.47E-01	8.68E-01	9.26E-01	9.88E-01	9.57E-01	8.74E-01	9.38E-01
GD	5	9.57E-01	1.78E-01	9.38E-01	9.38E-01	8.74E-01	9.26E-01	9.19E-01	8.30E-01	1.78E-01	2.92E-01
GD	6	1.78E-01	3.30E-01	1.78E-01	7.68E-01	5.80E-01	5.11E-01	9.80E-01	7.68E-01	3.50E-01	5.47E-01
GD	7	8.74E-01	3.00E-02	9.57E-01	1.86E-01	8.74E-01	9.57E-01	9.19E-01	8.74E-01	8.44E-02	1.78E-01
GD	8	9.57E-01	1.78E-01	5.11E-01	9.38E-01	8.30E-01	7.34E-01	9.38E-01	7.71E-01	2.00E-01	3.30E-01
GD	9	8.74E-01	8.30E-01	8.30E-01	8.74E-01	8.30E-01	9.57E-01	9.57E-01	9.38E-01	8.74E-01	9.57E-01
GD	10	8.30E-01	9.57E-01	9.38E-01	9.57E-01	8.30E-01	9.38E-01	8.30E-01	8.74E-01	9.77E-01	9.57E-01
GD	11	9.57E-01	8.30E-01	8.69E-01	9.80E-01	8.74E-01	7.68E-01	8.68E-01	9.57E-01	9.38E-01	9.26E-01
GD	12	2.81E-01	8.44E-02	7.09E-01	5.70E-01	7.34E-01	8.74E-01	7.68E-01	5.70E-01	2.81E-01	3.30E-01

Table D. 5. Linear mixed effects models comparing the effect of time and mortality group on TCR module score.

P values are reported for each main effect term. Additionally, p values are provided for the ANOVA test, which compares the variance explained by the model to that explained by the null model or the covariate-only model. All p values are reported to 3 significant figures. Survivors are used as the baseline.

Receptor	Module	Day	Simple Effects Tests					ANOVA
			MORTALITY LATE DEATH	MORTALITY EARLY DEATH	AGE	SEX.MALE	COMORBIDITIES CI	COV ONLY MODEL
Raw p values								
AB	11	Day1	1.55E-01	9.81E-01	9.19E-01	1.75E-01	4.20E-01	3.11E-01
		Day3	2.10E-01	8.34E-02	1.78E-02	3.02E-01	9.05E-01	5.49E-02
		Day5	4.29E-01	6.79E-01	5.84E-01	4.57E-01	9.36E-01	7.07E-01
	16	Day1	8.80E-01	2.08E-03	3.27E-02	9.16E-01	8.13E-01	4.93E-03
		Day3	6.98E-01	3.11E-03	7.15E-01	2.13E-01	3.10E-02	6.78E-03
		Day5	9.36E-01	7.88E-01	2.05E-01	6.95E-02	2.10E-01	9.64E-01
	7	Day1	8.61E-02	3.39E-04	7.95E-01	7.82E-01	4.60E-01	2.21E-05
		Day3	6.39E-01	8.90E-03	7.28E-01	2.10E-01	6.51E-01	1.75E-02
		Day5	9.99E-01	6.30E-01	7.70E-01	5.81E-01	9.69E-01	8.83E-01
GD	7	Day1	6.49E-01	7.44E-03	3.42E-01	8.15E-01	7.48E-01	2.32E-02
		Day3	7.66E-01	2.19E-01	2.76E-01	5.30E-01	3.64E-01	3.55E-01
		Day5	3.52E-01	6.85E-01	5.69E-01	3.41E-01	2.56E-01	6.33E-01
BH Adjusted p values								
AB	11	Day1	5.22E-01	9.99E-01	9.99E-01	5.23E-01	8.87E-01	7.00E-01
		Day3	5.23E-01	3.10E-01	1.07E-01	7.00E-01	9.99E-01	2.47E-01
		Day5	8.87E-01	9.99E-01	9.99E-01	8.87E-01	9.99E-01	9.99E-01
	16	Day1	9.99E-01	3.74E-02	1.61E-01	9.99E-01	9.99E-01	5.33E-02
		Day3	9.99E-01	4.20E-02	9.99E-01	5.23E-01	1.61E-01	6.10E-02
		Day5	9.99E-01	9.99E-01	5.23E-01	2.89E-01	5.23E-01	9.99E-01
	7	Day1	3.10E-01	9.15E-03	9.99E-01	9.99E-01	8.87E-01	1.19E-03
		Day3	9.99E-01	6.86E-02	9.99E-01	5.23E-01	9.99E-01	1.07E-01
		Day5	9.99E-01	9.99E-01	9.99E-01	9.99E-01	9.99E-01	9.99E-01
GD	7	Day1	8.11E-01	1.34E-01	6.55E-01	8.15E-01	8.11E-01	2.09E-01
		Day3	8.11E-01	6.55E-01	6.55E-01	8.11E-01	6.55E-01	6.55E-01
		Day5	6.55E-01	8.11E-01	8.11E-01	6.55E-01	6.55E-01	8.11E-01

Table D. 6. Simple effects tests comparing TCR module score between mortality groups at each time point.

P values are reported for each main effect term. Additionally, p values are provided for the ANOVA test, which compares the variance explained by the model to that explained by the covariate-only model. All p values are reported to 3 significant figures. Survivors are used as the baseline.

Receptor	Module	Mortality Group	Simple Effects Tests				ANOVA	
			TIMEPOINT	AGE	SEX.MALE	COMORBIDITIES.CI	COV ONLY MODEL	NULL MODEL
Raw p values								
AB	11	EARLY.DEATH	1.28E-01	7.40E-01	2.71E-01	8.96E-01	1.84E-01	6.21E-01
		LATE.DEATH	3.50E-01	5.42E-01	2.23E-01	4.05E-02	3.46E-01	1.99E-01
		SURVIVORS	2.70E-01	1.67E-02	1.41E-01	6.36E-02	2.67E-01	5.43E-02
	16	EARLY.DEATH	2.62E-01	8.36E-01	2.04E-02	6.35E-01	2.58E-01	7.86E-02
		LATE.DEATH	4.34E-01	4.58E-01	6.58E-01	8.51E-01	4.29E-01	7.67E-01
		SURVIVORS	6.78E-01	3.08E-03	5.91E-01	3.61E-02	6.76E-01	3.68E-02
	7	EARLY.DEATH	4.73E-01	5.56E-01	9.06E-01	9.61E-01	4.63E-01	9.27E-01
		LATE.DEATH	2.35E-03	8.83E-01	2.96E-01	2.18E-02	1.88E-03	2.15E-03
		SURVIVORS	4.45E-01	3.98E-02	8.66E-01	4.85E-01	4.43E-01	1.94E-01
GD	7	EARLY.DEATH	1.29E-01	9.29E-01	9.57E-01	9.01E-01	1.11E-01	6.09E-01
		LATE.DEATH	6.07E-01	9.45E-01	9.52E-02	6.16E-02	6.02E-01	4.64E-02
		SURVIVORS	5.31E-01	5.95E-01	3.87E-02	3.30E-01	5.30E-01	2.32E-01
BH Adjusted p values								
AB	11	EARLY.DEATH	4.61E-01	8.88E-01	5.85E-01	9.40E-01	5.66E-01	8.32E-01
		LATE.DEATH	6.74E-01	7.90E-01	5.85E-01	1.99E-01	6.74E-01	5.66E-01
		SURVIVORS	5.85E-01	1.68E-01	4.77E-01	2.64E-01	5.85E-01	2.44E-01
	16	EARLY.DEATH	5.85E-01	9.40E-01	1.68E-01	8.32E-01	5.85E-01	3.03E-01
		LATE.DEATH	7.28E-01	7.28E-01	8.32E-01	9.40E-01	7.28E-01	9.01E-01
		SURVIVORS	8.32E-01	4.16E-02	8.18E-01	1.99E-01	8.32E-01	1.99E-01
	7	EARLY.DEATH	7.28E-01	7.90E-01	9.40E-01	9.61E-01	7.28E-01	9.44E-01
		LATE.DEATH	4.16E-02	9.40E-01	6.15E-01	1.68E-01	4.16E-02	4.16E-02
		SURVIVORS	7.28E-01	1.99E-01	9.40E-01	7.28E-01	7.28E-01	5.66E-01
GD	7	EARLY.DEATH	3.87E-01	9.57E-01	9.57E-01	9.57E-01	3.87E-01	7.83E-01
		LATE.DEATH	7.83E-01	9.57E-01	3.87E-01	3.70E-01	7.83E-01	3.70E-01
		SURVIVORS	7.83E-01	7.83E-01	3.70E-01	7.43E-01	7.83E-01	5.96E-01

Table D. 7. Simple effects tests comparing the effect of time in each mortality group on TCR module score.

P values are reported for each main effect term. Additionally, p values are provided for the ANOVA test, which compares the variance explained by the model to that explained by the null model or the covariate-only model. All p values are reported to 3 significant figures.

AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	Receptor
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Module
1.59E-01	1.28E-01	9.74E-01	1.96E-01	9.88E-01	3.29E-01	5.54E-01	5.12E-01	3.12E-01	3.30E-01	4.02E-01	4.43E-01	1.17E-01	9.50E-01	5.09E-01	**Arterial pH**
5.41E-01	9.51E-01	1.66E-01	4.72E-01	4.62E-01	2.32E-01	6.09E-01	2.72E-01	3.56E-01	2.85E-01	4.28E-01	6.70E-01	9.76E-01	8.53E-01	4.43E-01	**Cardiogenic Pulmonary Oedema**
7.67E-01	5.78E-01	2.62E-02	9.12E-01	2.73E-01	9.07E-01	7.79E-01	7.78E-01	7.73E-01	8.01E-01	4.37E-01	1.43E-01	9.46E-01	8.79E-01	6.13E-01	**Haematocrit**
5.42E-01	6.02E-01	9.64E-01	6.77E-02	8.19E-01	8.36E-01	1.43E-01	2.54E-02	2.68E-02	6.72E-01	7.87E-01	8.01E-01	6.60E-01	3.01E-01	4.75E-01	**Infection: Number of ICU Acquired**
1.83E-01	2.39E-01	5.58E-01	3.89E-01	6.66E-02	1.73E-01	1.72E-01	4.07E-02	5.50E-02	6.40E-01	1.50E-01	1.50E-01	9.73E-01	1.39E-01	9.13E-01	**Infection: Ventilator Associated Pneumonia**
9.13E-01	6.95E-02	1.66E-01	1.40E-02	5.46E-01	1.27E-01	4.13E-01	4.29E-01	8.76E-01	2.85E-01	7.29E-01	3.40E-01	8.87E-01	2.78E-01	4.13E-01	**Lung Infiltrates: Diffuse Bilateral**
9.93E-01	5.54E-01	8.36E-01	5.34E-02	6.72E-01	9.99E-02	1.75E-01	3.60E-01	2.38E-01	3.68E-01	7.96E-01	1.53E-01	6.25E-01	3.81E-01	7.93E-01	**Lung Infiltrates: Lobar**
9.72E-01	1.41E-01	4.64E-01	3.86E-01	3.49E-01	4.21E-01	9.97E-01	9.54E-01	6.09E-01	1.25E-01	6.58E-01	8.04E-01	9.80E-01	7.46E-01	4.95E-01	**Lung Infiltrates: Localised**
6.70E-01	1.54E-02	9.30E-01	8.60E-01	6.98E-01	4.54E-01	2.66E-01	8.85E-01	9.02E-01	9.96E-01	6.51E-01	3.82E-01	6.79E-01	5.65E-02	4.90E-01	**Organism: Haemophilus Influenza**
3.03E-01	1.11E-01	3.80E-01	8.03E-01	4.20E-01	6.87E-01	5.04E-01	4.01E-01	4.68E-02	6.80E-01	5.05E-01	7.89E-01	7.49E-01	1.53E-01	6.33E-01	**Organism: Streptococcus Pneumonia**
8.79E-01	1.88E-02	4.18E-01	5.94E-01	9.74E-01	1.54E-02	9.23E-01	4.53E-01	6.47E-01	6.14E-01	7.41E-05	1.72E-02	2.86E-02	3.29E-03	6.70E-04	**Potassium: Highest**
9.43E-03	2.29E-01	4.04E-01	4.86E-01	9.01E-01	8.40E-02	1.49E-01	3.87E-01	6.50E-02	2.41E-01	4.97E-01	6.58E-01	1.83E-02	9.14E-01	7.41E-01	**Sodium: Highest**
8.18E-01	7.45E-01	6.01E-01	8.95E-01	6.47E-01	1.32E-01	4.71E-01	6.50E-01	7.84E-01	2.94E-01	4.23E-02	6.31E-03	1.12E-01	9.27E-01	3.34E-01	**Therapy: Corticosteroids**
5.35E-01	9.79E-01	1.33E-01	9.72E-01	7.31E-01	1.22E-01	3.97E-01	2.68E-01	2.43E-01	6.83E-01	6.92E-01	6.66E-01	7.96E-01	9.09E-01	9.97E-02	**Therapy: Days of Inotrope Pressor Support**
2.41E-01	4.74E-01	4.86E-01	3.01E-02	2.40E-01	2.05E-01	6.04E-01	7.03E-01	1.29E-02	1.59E-02	6.88E-01	9.47E-01	1.76E-01	4.64E-01	6.02E-01	**Therapy: Drotecogin Alfa**
2.74E-01	4.41E-01	3.60E-01	7.09E-01	9.83E-01	9.82E-01	6.71E-01	7.35E-01	6.11E-01	3.08E-01	1.82E-01	6.11E-02	5.82E-01	4.95E-01	2.88E-01	**Therapy: Mechanical Ventilation CPAP**
5.96E-01	9.67E-01	7.24E-01	7.39E-01	6.21E-01	8.92E-01	9.24E-01	8.92E-01	8.64E-01	3.14E-01	5.58E-01	2.52E-01	4.23E-01	5.92E-01	9.50E-01	**Therapy: Number of Sepsis Specific**
5.38E-01	3.01E-01	6.77E-01	4.24E-01	2.49E-01	5.59E-01	7.81E-01	6.18E-01	8.96E-02	1.67E-01	2.02E-01	9.73E-01	1.29E-02	5.02E-01	6.00E-02	**Therapy: Renal Replacement Therapy**
4.65E-01	2.53E-02	4.51E-01	2.91E-01	7.32E-01	2.87E-01	9.44E-01	3.36E-01	4.64E-01	9.20E-01	5.27E-01	4.85E-01	1.11E-01	2.30E-01	2.66E-01	**Total: Apache**
2.87E-01	3.00E-02	6.78E-01	9.33E-01	8.73E-01	1.09E-01	4.54E-01	2.00E-01	6.64E-02	6.10E-01	9.61E-01	8.93E-01	7.39E-02	5.86E-01	5.76E-01	**Total: SAPS**

GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	AB	AB	Receptor
12	11	10	9	8	7	6	5	4	3	2	1	17	16	Module
9.49E-01	2.50E-01	5.26E-01	4.14E-01	6.76E-02	4.22E-01	5.63E-01	3.02E-01	1.37E-01	6.92E-01	3.59E-02	5.20E-01	2.62E-02	1.51E-01	**Arterial pH**
2.22E-01	8.17E-01	7.97E-01	6.67E-01	6.66E-01	2.61E-01	6.46E-02	7.24E-02	3.21E-01	3.11E-01	2.13E-01	4.03E-01	9.67E-01	5.26E-01	**Cardiogenic Pulmonary Oedema**
3.11E-01	2.61E-01	3.05E-01	4.95E-01	9.32E-01	9.45E-01	3.57E-01	3.24E-01	9.73E-01	2.66E-01	4.49E-01	4.64E-01	3.35E-01	8.98E-01	**Haematocrit**
5.25E-01	6.06E-02	5.10E-03	8.37E-01	6.98E-01	9.65E-02	4.49E-01	6.10E-01	4.40E-01	3.09E-02	8.13E-01	5.63E-01	8.56E-01	4.38E-01	**Infection: Number of ICU Acquired**
3.14E-01	3.49E-02	1.06E-02	8.07E-01	5.14E-01	4.41E-02	4.97E-01	7.37E-01	2.21E-01	1.06E-01	2.05E-01	2.07E-01	5.93E-01	5.40E-01	**Infection: Ventilator Associated Pneumonia**
9.18E-01	3.20E-01	5.39E-01	6.53E-01	1.20E-01	6.51E-01	6.74E-01	3.39E-01	6.37E-01	7.33E-02	9.39E-01	5.38E-01	1.85E-01	8.46E-01	**Lung Infiltrates: Diffuse Bilateral**
6.47E-01	3.53E-01	5.54E-01	3.23E-01	3.72E-01	4.63E-01	9.05E-01	7.68E-02	7.37E-01	7.59E-03	6.53E-01	8.49E-01	1.48E-01	6.44E-01	**Lung Infiltrates: Lobar**
5.32E-01	3.73E-01	3.98E-01	8.99E-01	2.73E-01	1.84E-01	8.21E-01	7.62E-01	9.80E-01	2.74E-01	5.39E-01	6.85E-01	9.66E-01	5.62E-01	**Lung Infiltrates: Localised**
4.33E-01	5.67E-01	4.06E-01	9.53E-01	9.29E-01	2.28E-01	9.06E-01	3.01E-01	2.66E-01	8.10E-01	3.41E-01	6.86E-01	6.50E-01	6.05E-01	**Organism: Haemophilus Influenza**
5.24E-01	3.94E-04	1.12E-03	6.11E-01	9.60E-01	9.38E-01	8.67E-01	3.97E-01	2.76E-01	2.11E-02	8.11E-03	1.43E-02	1.60E-02	6.57E-01	**Organism: Streptococcus Pneumonia**
5.71E-01	9.22E-01	8.50E-01	8.29E-01	3.10E-01	2.11E-01	5.29E-01	2.02E-01	7.79E-01	9.48E-01	3.58E-01	9.04E-01	5.99E-01	3.49E-01	**Potassium: Highest**
4.28E-02	1.76E-01	1.60E-01	1.92E-01	3.11E-01	9.69E-01	1.57E-01	7.94E-01	4.36E-01	8.25E-01	2.48E-01	1.54E-01	2.34E-01	1.58E-01	**Sodium: Highest**
8.74E-01	2.77E-01	9.21E-01	6.78E-01	5.36E-01	3.30E-02	8.86E-01	2.77E-01	3.92E-01	1.34E-01	3.06E-02	9.70E-01	5.44E-02	3.65E-01	**Therapy: Corticosteroids**
7.49E-01	8.76E-01	6.56E-01	8.97E-01	6.04E-01	5.66E-01	2.91E-01	8.72E-01	5.23E-01	5.72E-01	5.36E-01	2.80E-01	6.50E-01	6.64E-01	**Therapy: Days of Inotrope Pressor Support**
3.65E-01	2.25E-01	1.70E-01	7.04E-01	2.17E-01	3.16E-01	2.22E-01	5.60E-01	3.80E-01	8.82E-01	5.30E-01	1.05E-01	6.56E-01	7.86E-01	**Therapy: Drotecogin Alfa**
4.31E-01	2.36E-01	2.48E-01	4.60E-01	9.67E-01	4.25E-01	6.56E-01	8.32E-01	8.86E-01	4.19E-01	6.20E-01	3.69E-01	7.18E-01	3.11E-01	**Therapy: Mechanical Ventilation CPAP**
9.27E-01	4.06E-02	8.26E-02	9.17E-01	4.91E-01	3.38E-01	6.33E-02	9.78E-01	8.12E-01	2.27E-02	1.84E-01	3.12E-01	2.74E-02	1.63E-01	**Therapy: Number of Sepsis Specific**
1.06E-01	2.38E-01	1.05E-01	9.05E-01	4.80E-01	9.73E-01	8.19E-01	2.33E-01	6.25E-01	8.73E-01	4.40E-01	1.88E-01	9.51E-01	4.04E-01	**Therapy: Renal Replacement Therapy**
3.33E-01	6.86E-01	6.66E-01	8.43E-01	1.94E-01	4.16E-01	3.71E-01	2.66E-01	3.89E-01	5.94E-01	4.23E-01	5.29E-01	3.90E-01	9.53E-01	**Total: Apache**
1.26E-01	6.24E-01	8.69E-01	8.61E-01	5.69E-02	7.19E-01	2.11E-01	1.35E-01	3.87E-01	6.92E-01	2.61E-01	3.88E-01	3.04E-01	2.58E-01	**Total: SAPS**

Table D. 8. Linear mixed effects models investigating the effect of each clinical variable measured once per ICU admission on day 1 TCRAB module score.

P values are reported for the main effect term, to 3 significant figures.

AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	AB	Receptor
17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Module
4.32E-01	5.45E-01	9.38E-01	5.95E-03	3.21E-01	1.74E-01	4.15E-01	1.93E-01	9.23E-02	2.06E-01	4.03E-01	7.30E-01	8.87E-01	1.33E-01	1.83E-01	5.30E-01	1.90E-01	Bicarbonate
7.68E-01	1.23E-01	7.98E-01	9.91E-02	3.52E-01	6.96E-02	8.87E-02	5.78E-01	4.85E-01	7.53E-01	8.15E-01	2.00E-01	5.40E-01	6.23E-01	6.54E-07	7.34E-01	4.56E-01	Bilirubin: Highest
6.56E-01	5.98E-01	3.71E-01	3.36E-01	1.37E-01	9.71E-01	2.18E-01	8.07E-01	5.67E-01	4.05E-01	1.16E-01	1.40E-01	2.65E-01	3.22E-01	3.68E-01	7.63E-01	1.19E-01	Creatine: Highest
3.31E-01	2.09E-01	2.04E-01	4.92E-01	2.42E-01	1.64E-01	9.85E-01	8.03E-01	4.46E-01	7.19E-01	2.60E-01	1.70E-01	6.86E-01	9.13E-01	2.68E-01	7.66E-01	8.30E-01	GCS: Total
8.05E-01	8.66E-01	3.73E-01	6.75E-01	6.15E-01	4.56E-01	6.35E-01	1.76E-01	4.06E-01	6.98E-02	3.63E-01	8.04E-01	9.95E-01	6.42E-01	5.74E-01	3.78E-01	3.31E-01	Heart Rate: Highest
4.80E-01	2.62E-01	6.67E-01	8.40E-02	2.05E-01	4.97E-01	4.07E-01	7.23E-02	4.28E-01	4.10E-01	8.82E-01	1.01E-01	6.25E-01	4.57E-01	6.52E-01	1.05E-01	9.54E-01	Hypertension
7.72E-01	4.21E-01	5.89E-01	3.48E-01	3.89E-01	2.30E-01	9.02E-01	1.56E-01	5.89E-01	3.60E-01	3.02E-01	5.39E-02	9.83E-01	5.90E-01	9.73E-02	4.93E-01	8.51E-01	Leukocytes: Lymphocytes
7.63E-01	5.76E-01	5.96E-01	9.13E-01	9.55E-01	6.08E-02	1.59E-01	1.32E-01	3.86E-02	6.69E-01	6.68E-01	9.67E-01	4.38E-01	5.90E-01	7.53E-01	6.60E-01	3.25E-01	Leukocytes: Mononuclear
2.80E-01	2.61E-03	1.29E-01	2.35E-01	8.59E-01	3.21E-01	4.46E-01	3.96E-01	1.24E-02	9.94E-02	9.54E-01	5.98E-01	2.87E-01	2.43E-01	9.49E-01	4.55E-01	2.50E-01	Leukocytes: Polymorphonu
2.05E-01	4.32E-01	3.83E-01	9.34E-01	2.51E-01	2.12E-01	7.62E-01	4.21E-01	3.20E-01	6.12E-01	3.62E-01	5.62E-01	1.77E-01	2.89E-01	1.03E-01	5.57E-01	2.98E-01	Mean Arterial Pressure:
6.77E-01	6.08E-01	8.86E-01	8.41E-01	9.16E-01	7.43E-01	9.53E-01	3.88E-02	7.61E-01	1.83E-01	8.09E-01	2.26E-01	9.35E-01	8.10E-01	9.96E-01	8.75E-01	7.10E-01	MPa02
1.86E-02	9.21E-01	3.09E-05	7.66E-01	7.40E-01	9.23E-01	3.14E-01	6.19E-01	7.84E-01	9.57E-01	2.14E-01	8.15E-01	3.80E-01	1.99E-01	9.46E-01	9.03E-01	6.44E-01	PC02
1.35E-02	4.89E-01	5.39E-01	7.55E-01	2.74E-01	7.72E-01	1.21E-01	8.23E-01	1.65E-01	9.62E-01	9.64E-03	5.21E-01	4.63E-01	4.02E-02	1.10E-01	2.08E-01	5.08E-01	Platelets: Lowest
8.60E-01	8.19E-01	2.05E-01	9.27E-01	6.93E-01	6.31E-01	2.92E-01	7.34E-01	6.51E-01	5.33E-01	9.97E-01	1.03E-01	7.64E-01	7.38E-01	2.02E-01	8.21E-01	6.36E-01	Respiratory
3.06E-01	2.93E-01	1.09E-01	4.62E-01	4.05E-01	4.23E-01	6.32E-01	7.20E-02	9.68E-01	6.62E-01	1.88E-01	7.04E-01	8.11E-01	7.13E-01	6.57E-01	8.87E-01	4.00E-01	Systolic Blood Pressure:
5.92E-01	3.31E-01	5.95E-01	4.29E-01	4.01E-01	7.51E-01	3.83E-01	9.22E-01	2.53E-01	6.77E-01	8.59E-01	8.99E-01	3.46E-01	3.16E-01	2.23E-01	3.65E-01	2.36E-01	Temperature: Highest
7.43E-01	1.04E-01	9.02E-01	2.90E-01	2.30E-01	2.12E-01	7.30E-02	3.17E-01	2.78E-01	6.31E-01	1.07E-01	9.56E-01	8.66E-02	1.26E-02	8.52E-03	8.50E-01	3.34E-01	Total: SOFA
7.55E-01	8.08E-01	8.71E-01	7.82E-01	1.36E-01	5.65E-02	7.03E-01	2.99E-01	5.37E-06	1.02E-02	6.99E-01	6.74E-01	2.85E-02	3.05E-01	3.83E-01	9.96E-01	3.43E-02	Urea: Highest
1.39E-01	5.35E-03	5.89E-01	1.89E-01	3.27E-01	9.97E-01	4.18E-01	1.76E-01	4.59E-05	1.79E-02	5.45E-01	5.70E-01	2.18E-01	4.16E-01	2.57E-01	2.86E-01	7.24E-02	WCC: Highest

Receptor	Module	Bicarbonate	Bilirubin: Highest	Creatine: Highest	GCS: Total	Heart Rate: Highest	Hypertension	Leukocytes: Lymphocytes	Leukocytes: Mononuclear	Leukocytes: Polymorphonu	Mean Arterial Pressure:	MPa02	PC02	Platelets: Lowest	Respiratory	Systolic Blood Pressure:	Temperature: Highest	Total: SOFA	Urea: Highest	WCC: Highest
GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD	GD
12	11	10	9	8	7	6	5	4	3	2	1									
4.24E-01	3.52E-01	2.97E-01	7.76E-01	1.01E-01	9.94E-01	6.53E-02	1.25E-01	1.84E-01	2.29E-01	7.52E-01	8.40E-01									
7.60E-01	5.72E-01	6.74E-01	2.49E-01	7.54E-01	3.64E-01	6.52E-01	2.94E-02	3.70E-01	3.57E-01	2.24E-01	9.29E-01									
1.79E-01	3.01E-02	2.06E-01	6.94E-01	7.73E-02	7.84E-01	3.28E-01	1.51E-01	1.94E-01	6.51E-01	6.57E-01	1.34E-02									
2.50E-01	6.92E-01	3.72E-01	8.32E-01	9.42E-02	1.74E-01	2.66E-01	6.55E-01	3.74E-01	2.23E-01	7.26E-01	2.33E-01									
6.87E-01	5.13E-01	5.83E-01	4.52E-01	1.07E-01	8.94E-01	7.78E-01	3.82E-01	5.10E-01	7.08E-01	5.69E-01	6.59E-01									
4.54E-01	8.69E-01	9.78E-01	4.52E-01	1.93E-01	2.69E-01	3.20E-01	8.86E-01	1.19E-01	9.82E-01	2.88E-01	8.50E-01									
1.43E-01	4.04E-01	9.55E-01	1.99E-02	7.95E-01	8.13E-01	5.80E-01	4.53E-03	5.38E-01	3.06E-01	6.51E-02	3.77E-01									
5.15E-01	8.99E-01	2.63E-01	2.89E-01	6.64E-01	8.18E-01	9.98E-01	8.37E-03	5.84E-01	3.59E-01	2.80E-01	8.36E-01									
7.80E-01	6.06E-02	1.29E-01	9.77E-01	8.39E-01	9.23E-01	3.47E-01	1.32E-01	1.31E-01	7.55E-01	4.38E-03	1.45E-01									
9.02E-01	4.72E-01	2.35E-01	6.59E-01	2.61E-01	7.17E-01	3.90E-01	3.56E-01	3.55E-01	3.78E-01	7.06E-01	5.04E-01									
4.26E-01	2.75E-01	8.66E-01	5.60E-01	6.54E-01	9.20E-01	4.30E-01	6.38E-01	7.69E-01	5.98E-01	2.66E-01	3.54E-01									
4.83E-01	5.78E-01	9.11E-01	8.38E-01	4.79E-01	5.98E-01	7.15E-01	4.24E-01	1.83E-01	7.99E-01	6.54E-01	2.37E-01									
2.18E-01	4.87E-01	9.74E-02	2.47E-01	3.12E-01	1.44E-01	1.70E-01	1.22E-01	7.36E-02	6.60E-01	3.86E-01	9.16E-01									
5.77E-01	4.74E-01	9.37E-01	6.30E-01	1.10E-02	2.83E-01	9.44E-01	3.80E-01	2.64E-02	5.48E-01	7.40E-01	3.93E-01									
2.99E-01	7.61E-02	3.43E-02	6.92E-01	3.93E-01	8.53E-01	4.05E-01	6.56E-01	1.08E-01	7.77E-01	5.84E-01	1.60E-01									
2.24E-01	7.32E-01	7.77E-01	2.41E-01	5.78E-01	5.39E-01	1.88E-01	2.85E-01	1.12E-01	8.73E-01	2.61E-01	7.55E-01									
9.82E-01	2.44E-01	6.71E-01	7.16E-01	2.67E-01	6.37E-01	8.81E-02	2.20E-01	7.61E-02	9.43E-01	4.12E-01	3.31E-01									
5.91E-02	3.51E-01	2.87E-01	6.57E-02	3.01E-01	4.59E-01	2.15E-01	2.51E-01	7.21E-01	8.42E-01	6.39E-01	4.93E-01									
8.00E-01	8.37E-03	2.99E-02	2.58E-01	6.92E-01	7.53E-01	6.29E-01	1.95E-02	5.87E-01	7.54E-01	1.36E-02	4.78E-02									

Table D. 9. Linear mixed effects models investigating the effect of each longitudinal clinical variable on TCRAB and TCRGD module score.

P values are reported for the main effect term, to 3 significant figures.

Receptor	Module	Linear Mixed Effects Models					ANOVA	
		TIME	SRSq	AGE	SEX MALE	CI	COV ONLY MODEL	NULL MODEL
Raw p values								
AB	1	6.74E-01	1.00E+00	9.02E-01	9.05E-01	2.86E-01	8.93E-01	8.68E-01
AB	2	6.46E-01	7.05E-01	1.28E-01	8.82E-02	1.75E-01	8.81E-01	1.94E-02
AB	3	3.22E-01	6.96E-01	1.57E-01	5.93E-01	8.67E-01	3.89E-01	3.37E-01
AB	4	5.05E-01	2.19E-01	9.95E-01	5.71E-01	7.95E-01	4.82E-01	8.56E-01
AB	5	6.26E-01	9.36E-01	6.25E-01	8.66E-01	3.03E-01	8.75E-01	9.34E-01
AB	6	2.63E-01	3.73E-01	5.28E-01	3.32E-01	9.34E-01	4.89E-01	7.15E-01
AB	7	5.43E-01	4.11E-01	9.84E-01	6.53E-01	9.41E-01	6.99E-01	9.69E-01
AB	8	6.20E-05	7.63E-01	1.37E-01	2.51E-01	6.94E-01	2.04E-05	8.93E-05
AB	9	3.50E-01	4.80E-03	2.03E-01	5.81E-01	4.19E-01	1.80E-02	8.03E-02
AB	10	3.21E-01	5.69E-01	1.06E-01	1.66E-01	5.63E-03	5.94E-01	3.79E-02
AB	11	9.97E-01	5.66E-01	2.90E-01	7.74E-01	5.55E-01	8.05E-01	8.83E-01
AB	12	7.99E-01	8.54E-01	8.95E-01	1.56E-01	6.53E-01	9.16E-01	7.68E-01
AB	13	9.60E-01	8.63E-01	1.04E-01	9.42E-01	1.05E-01	9.85E-01	6.52E-01
AB	14	8.08E-01	5.36E-01	1.90E-01	1.54E-02	2.49E-02	8.29E-01	1.00E-03
AB	15	4.87E-01	4.09E-01	9.51E-01	1.12E-01	9.05E-01	6.64E-01	6.35E-01
AB	16	8.47E-02	7.60E-01	6.42E-01	2.72E-02	5.78E-01	1.97E-01	8.52E-02
AB	17	2.47E-01	3.91E-01	4.82E-01	1.44E-02	1.52E-01	1.35E-01	5.60E-02
GD	1	1.67E-01	7.72E-01	5.04E-01	2.97E-01	9.12E-01	3.34E-01	4.96E-01
GD	2	1.03E-01	8.32E-01	2.72E-01	7.80E-01	6.12E-01	1.13E-01	1.65E-01
GD	3	2.60E-01	5.05E-01	4.48E-01	9.40E-01	9.10E-01	5.21E-01	7.82E-01
GD	4	1.44E-01	5.53E-01	6.10E-01	9.41E-01	7.44E-01	3.52E-01	7.02E-01
GD	5	4.31E-01	4.30E-01	5.67E-01	6.13E-01	1.94E-01	6.64E-01	7.59E-01
GD	6	5.60E-01	1.15E-01	1.65E-01	8.93E-01	2.68E-01	8.06E-02	1.97E-01
GD	7	6.51E-02	2.75E-01	9.08E-01	5.30E-01	3.81E-01	1.76E-01	4.23E-01
GD	8	2.08E-01	9.52E-01	3.18E-01	9.25E-01	1.86E-01	3.82E-01	5.66E-01
GD	9	5.14E-01	1.97E-01	7.96E-01	9.98E-01	8.95E-01	4.37E-01	8.72E-01
GD	10	5.89E-01	9.15E-01	6.67E-01	4.04E-01	6.06E-01	8.53E-01	7.94E-01
GD	11	9.37E-01	8.10E-01	2.86E-01	5.19E-01	7.54E-01	9.72E-01	6.56E-01
GD	12	7.69E-01	2.17E-01	6.21E-01	4.01E-01	1.10E-01	2.68E-01	3.19E-01
BH Adjusted p Values								
AB	1	1.00E+00	1.00E+00	1.00E+00	1.00E+00	8.86E-01	1.00E+00	1.00E+00
AB	2	1.00E+00	1.00E+00	6.50E-01	5.83E-01	6.96E-01	1.00E+00	2.30E-01
AB	3	9.12E-01	1.00E+00	6.66E-01	1.00E+00	1.00E+00	9.70E-01	9.12E-01
AB	4	1.00E+00	7.65E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
AB	5	1.00E+00	1.00E+00	1.00E+00	1.00E+00	9.00E-01	1.00E+00	1.00E+00
AB	6	8.47E-01	9.65E-01	1.00E+00	9.12E-01	1.00E+00	1.00E+00	1.00E+00
AB	7	1.00E+00	9.77E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
AB	8	3.54E-03	1.00E+00	6.50E-01	8.28E-01	1.00E+00	2.43E-03	3.54E-03
AB	9	9.26E-01	1.12E-01	7.34E-01	1.00E+00	9.77E-01	2.30E-01	5.83E-01
AB	10	9.12E-01	1.00E+00	5.99E-01	6.82E-01	1.12E-01	1.00E+00	3.47E-01
AB	11	1.00E+00	1.00E+00	8.86E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00
AB	12	1.00E+00	1.00E+00	1.00E+00	6.66E-01	1.00E+00	1.00E+00	1.00E+00
AB	13	1.00E+00	1.00E+00	5.99E-01	1.00E+00	5.99E-01	1.00E+00	1.00E+00
AB	14	1.00E+00	1.00E+00	7.30E-01	2.28E-01	2.69E-01	1.00E+00	2.98E-02
AB	15	1.00E+00	9.77E-01	1.00E+00	6.05E-01	1.00E+00	1.00E+00	1.00E+00
AB	16	5.83E-01	1.00E+00	1.00E+00	2.69E-01	1.00E+00	7.31E-01	5.83E-01
AB	17	8.28E-01	9.70E-01	1.00E+00	2.28E-01	6.66E-01	6.50E-01	4.76E-01
GD	1	9.67E-01	9.76E-01	9.67E-01	9.67E-01	9.76E-01	9.67E-01	9.67E-01
GD	2	9.67E-01	9.76E-01	9.67E-01	9.76E-01	9.67E-01	9.67E-01	9.67E-01
GD	3	9.67E-01	9.67E-01	9.67E-01	9.76E-01	9.76E-01	9.67E-01	9.76E-01
GD	4	9.67E-01	9.67E-01	9.67E-01	9.76E-01	9.76E-01	9.67E-01	9.76E-01
GD	5	9.67E-01	9.67E-01	9.67E-01	9.67E-01	9.67E-01	9.76E-01	9.76E-01
GD	6	9.67E-01	9.67E-01	9.67E-01	9.76E-01	9.67E-01	9.67E-01	9.67E-01
GD	7	9.67E-01	9.67E-01	9.76E-01	9.67E-01	9.67E-01	9.67E-01	9.67E-01
GD	8	9.67E-01	9.76E-01	9.67E-01	9.76E-01	9.67E-01	9.67E-01	9.67E-01
GD	9	9.67E-01	9.67E-01	9.76E-01	9.98E-01	9.76E-01	9.67E-01	9.76E-01
GD	10	9.67E-01	9.76E-01	9.76E-01	9.67E-01	9.67E-01	9.76E-01	9.76E-01
GD	11	9.76E-01	9.76E-01	9.67E-01	9.67E-01	9.76E-01	9.84E-01	9.76E-01
GD	12	9.76E-01	9.67E-01	9.67E-01	9.67E-01	9.67E-01	9.67E-01	9.67E-01

Table D. 10. Linear mixed effects models exploring the effect of SRSq score and time on TCRAB module score in sepsis patients only.

P values are reported for each main effect term. Additionally, p values are provided for the ANOVA test, which compares the variance explained by the model to that explained by the null model or the covariate-only model. All p values are reported to 3 significant figures.

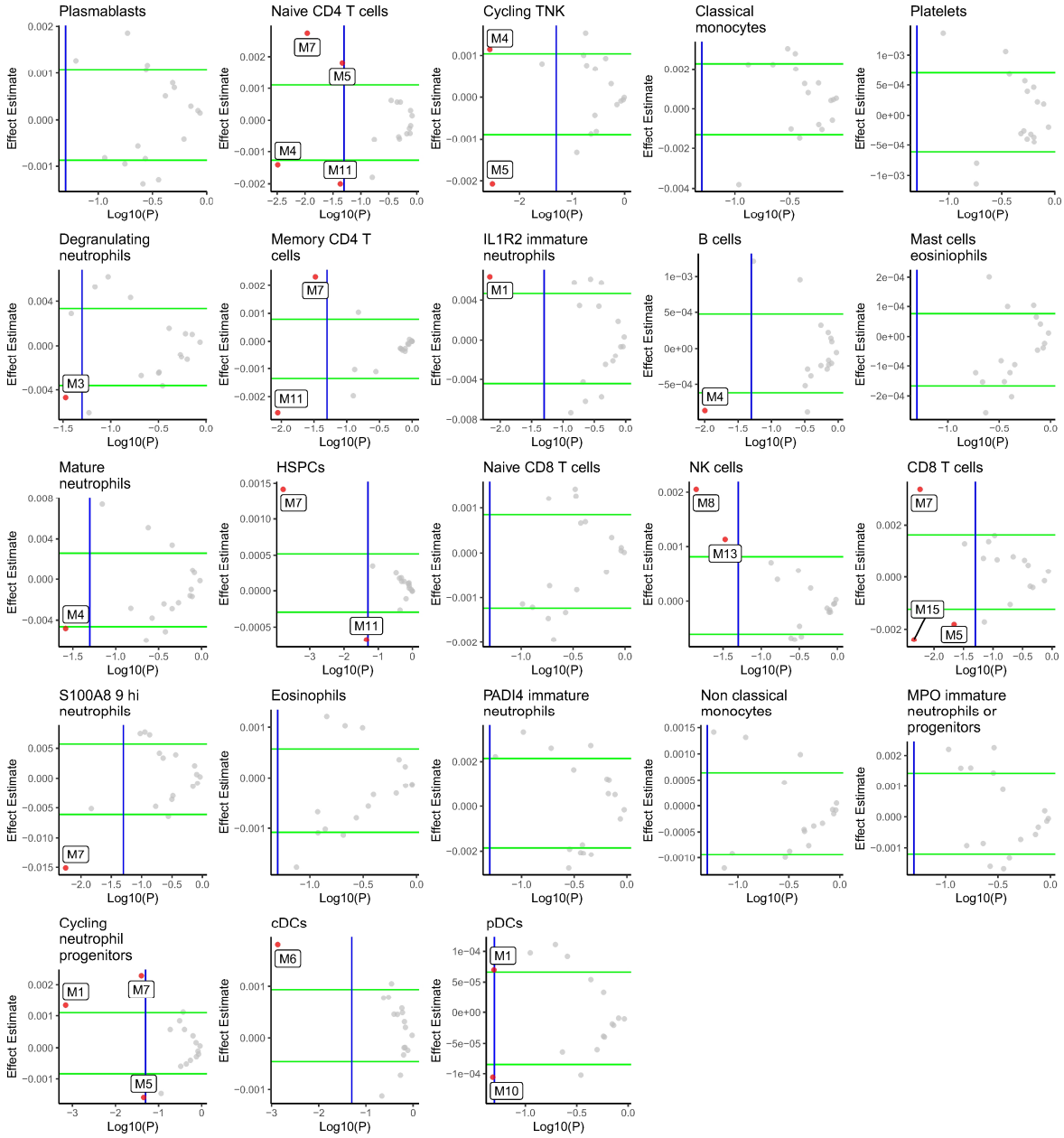


Figure D. 12. Association between TCRAB module score and CIBERSORTx estimated cell fraction using module score as the independent variable and cell fraction as the dependent variable.

Estimated effect sizes on cell fraction for each module and their corresponding p values are plotted. Log10(p=0.05) is shown in blue and the mean ± 1 SD is shown in green. Modules which meet these thresholds are annotated.

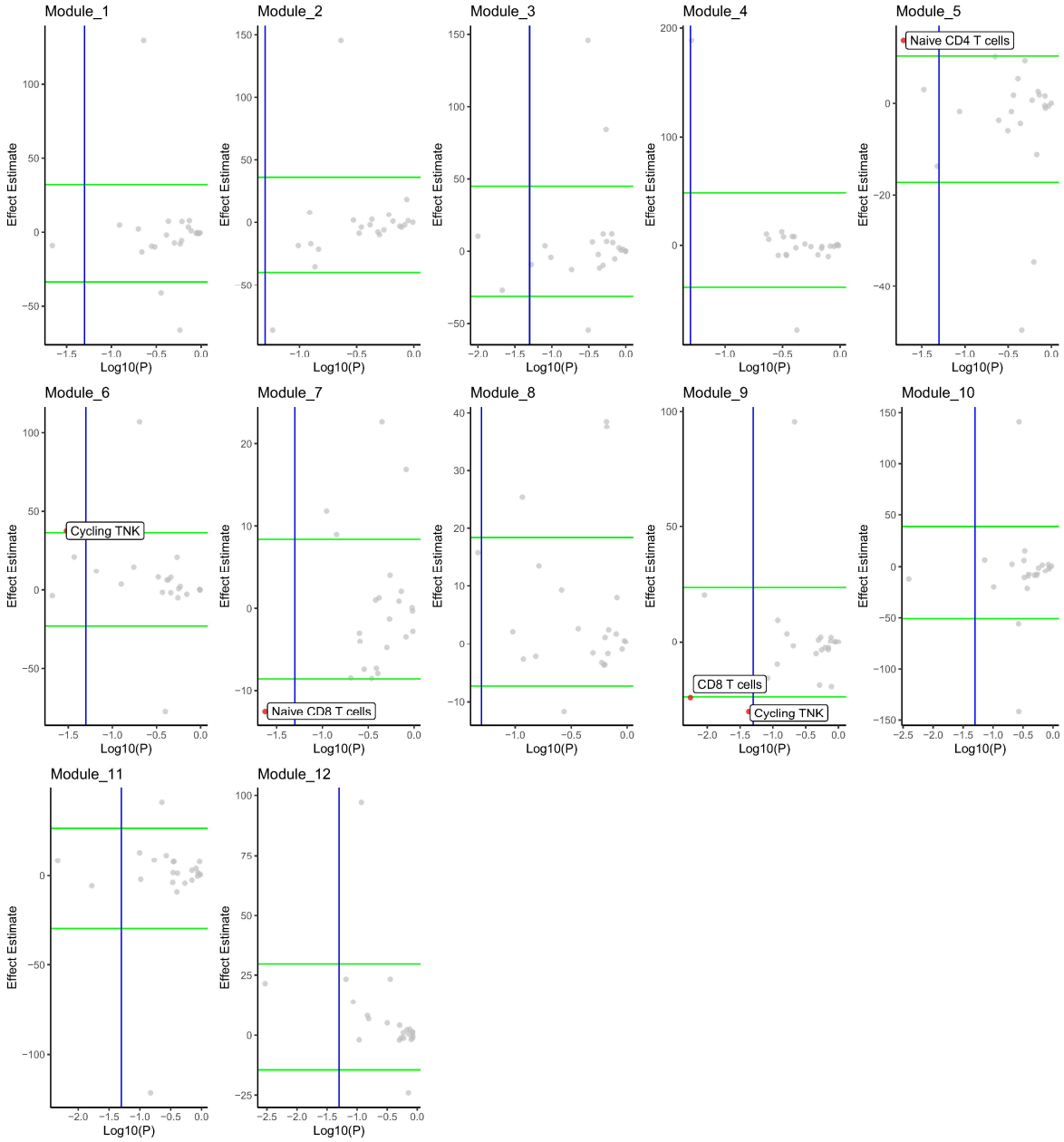


Figure D. 13. Association between CIBERSORTx estimated cell fraction and TCRGD module score.

Module score was the dependent variable and cell fraction was the independent variable. Estimated effect sizes on module score for each cell fraction and their corresponding p values are plotted. Log10(p=0.05) is shown in blue and the mean ± 1 SD is shown in green. Cell types which meet these thresholds are annotated.

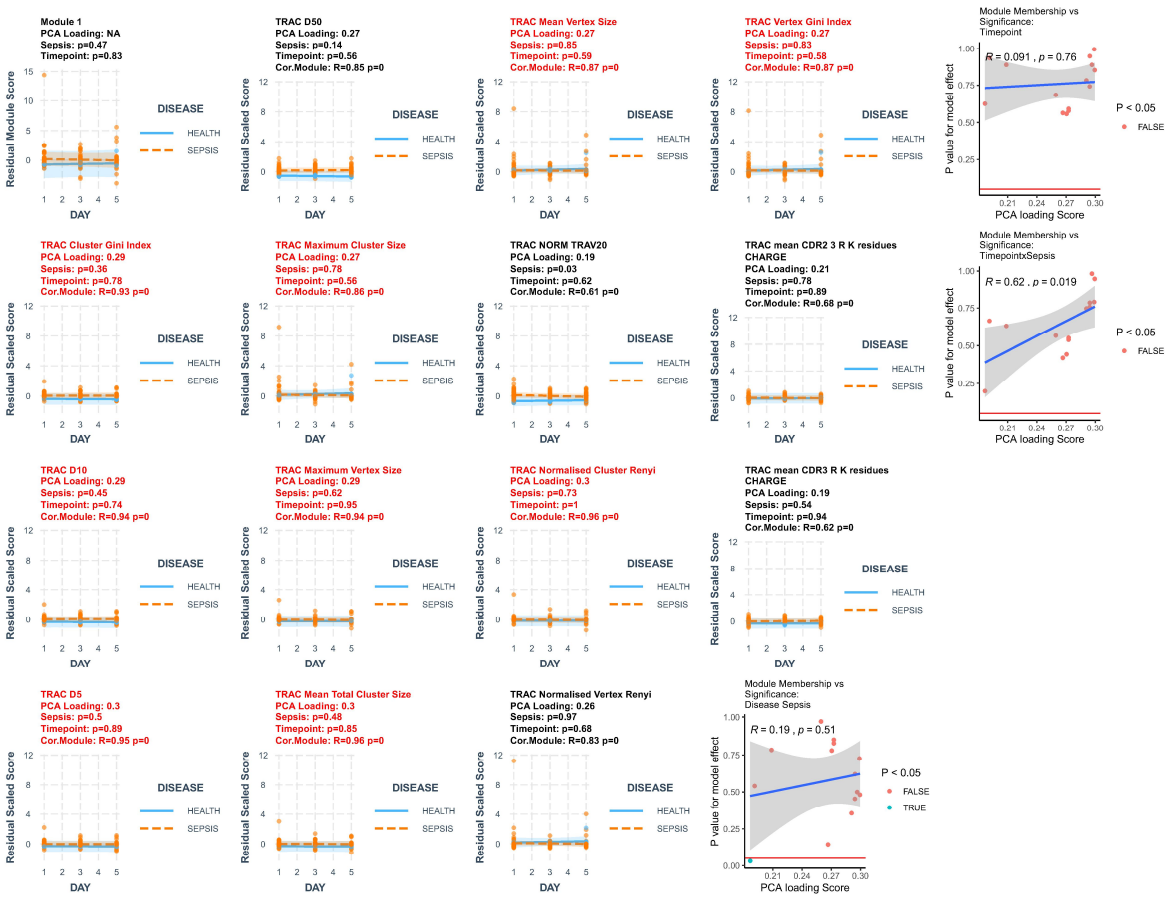


Figure D. 14. Partial residual plots showing the non-significant increase in TCRAB module 1 score in sepsis compared to health, and for the constituent features.

Top left shows the partial residual plot for TCRAB module 1 using the D*TASC linear mixed effect model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 1 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). 95% confidence intervals are shown.

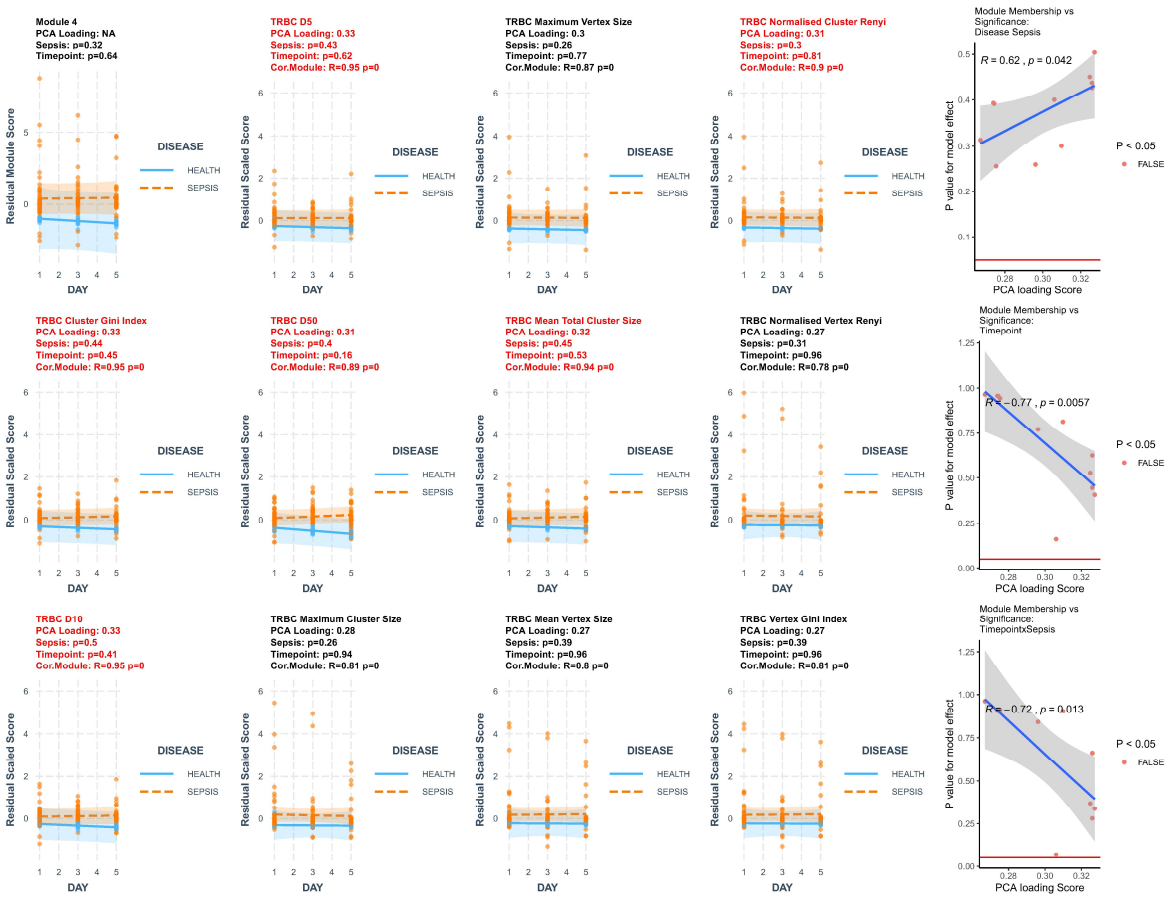


Figure D. 15. Partial residual plots showing the non-significant increase in TCRAB module 4 score in sepsis compared to health, and for the constituent features.

Top left shows the partial residual plot for TCRAB module 4 using the D*TASC linear mixed effect model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 4 cluster, modelled using the same D*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are three scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (disease sepsis, time point and disease time point interaction). 95% confidence intervals are shown.

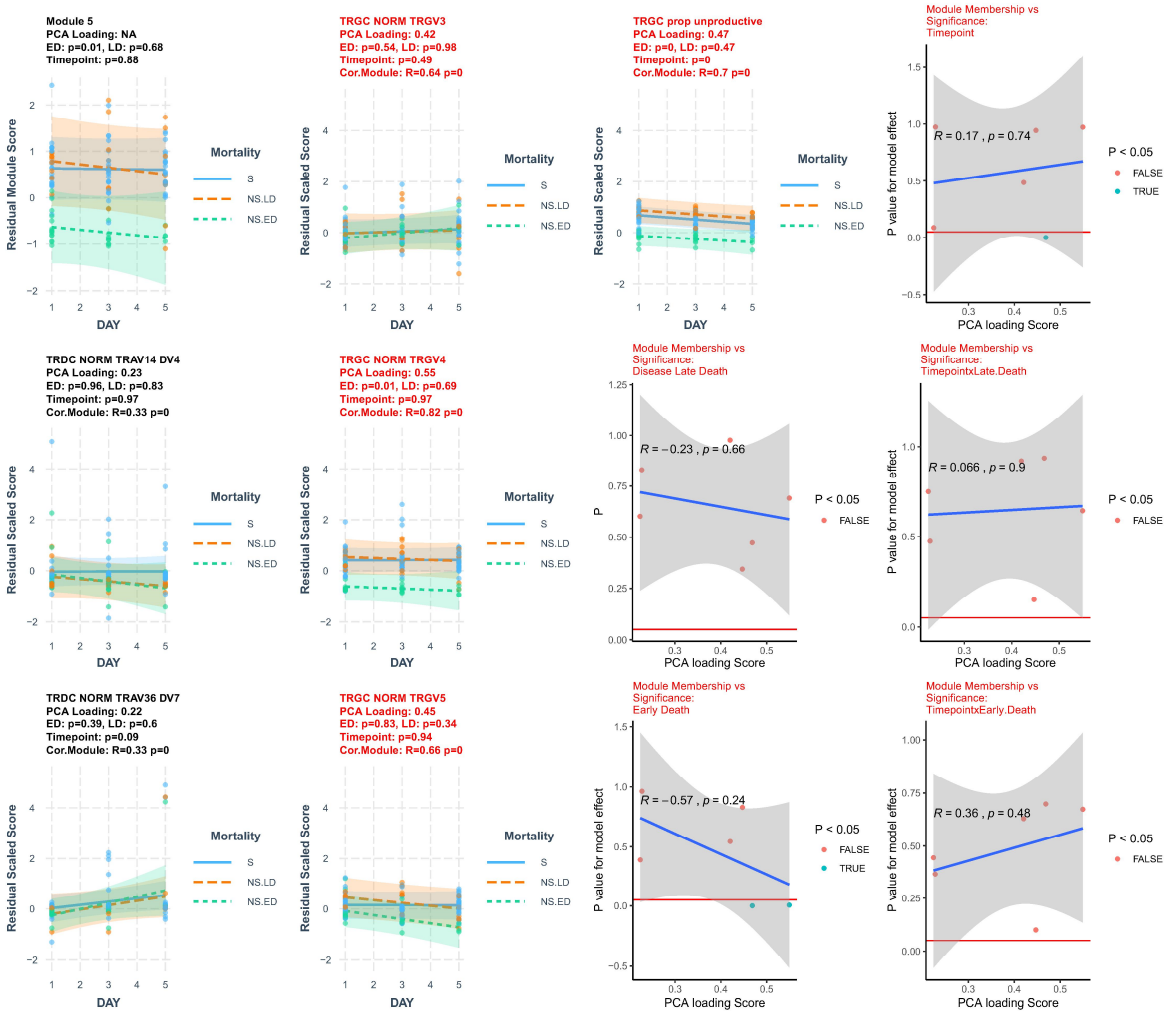


Figure D. 16. Partial residual plots showing the decrease in TCRGD module 5 score in sepsis non-survivors compared to survivors, corresponding to changes in T cell clonal expansion and diversification.

Top left shows the partial residual plot for TCRGD module 5 using the M*TASC linear mixed effect model and controlling for covariates. The remaining plots are partial residual plots for each scaled feature in the module 5 cluster, modelled using the same M*TASC model. The PC1 loading score is recorded above each plot, red text indicates a large loading score, highlighting a feature of interest. P values for the sepsis and time point main effects are also reported. The correlation (PCC, R) between the feature and the module eigenvector is also reported. On the far right are five scatter plots, investigating the relationship between the PC1 loading score and the p value for different main effect terms (sepsis mortality, time point and mortality time point interaction). 95% confidence intervals are shown.

References

ABDALLAH, F., COINDRE, S., GARDET, M., MEURISSE, F., NAJI, A., SUGANUMA, N., ABI-RACHED, L., LAMBOTTE, O. & FAVIER, B. 2021. Leukocyte Immunoglobulin-Like Receptors in Regulating the Immune Response in Infectious Diseases: A Window of Opportunity to Pathogen Persistence and a Sound Target in Therapeutics. *Front Immunol,* 12**,** 717998.

ABE, T., OGURA, H., KUSHIMOTO, S., SHIRAISHI, A., SUGIYAMA, T., DESHPANDE, G. A., UCHIDA, M., NAGATA, I., SAITOH, D., FUJISHIMA, S., MAYUMI, T., HIFUMI, T., SHIINO, Y., NAKADA, T. A., TARUI, T., OTOMO, Y., OKAMOTO, K., UMEMURA, Y., KOTANI, J., SAKAMOTO, Y., SASAKI, J., SHIRAISHI, S. I., TAKUMA, K., TSURUTA, R., HAGIWARA, A., YAMAKAWA, K., MASUNO, T., TAKEYAMA, N., YAMASHITA, N., IKEDA, H., UEYAMA, M., FUJIMI, S., GANDO, S. & GROUP, J. F. 2019. Variations in infection sites and mortality rates among patients in intensive care units with severe sepsis and septic shock in Japan. *J Intensive Care,* 7**,** 28.

AGATHANGELIDIS, A., CHATZIDIMITRIOU, A., CHATZIKONSTANTINOU, T., TRESOLDI, C., DAVIS, Z., GIUDICELLI, V., KOSSIDA, S., BELESSI, C., ROSENQUIST, R., GHIA, P., LANGERAK, A. W., DAVI, F. & STAMATOPOULOS, K. 2022. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: the 2022 update of the recommendations by ERIC, the European Research Initiative on CLL. *Leukemia,* 36**,** 1961-1968.

AHERN, D. J., AI, Z., AINSWORTH, M., ALLAN, C., ALLCOCK, A., ANGUS, B., ANSARI, M. A., ARANCIBIA-CÁRCAMO, C. V., ASCHENBRENNER, D., ATTAR, M., BAILLIE, J. K., BARNES, E., BASHFORD-ROGERS, R., BASHYAL, A., BEER, S., BERRIDGE, G., BEVERIDGE, A., BIBI, S., BICANIC, T., BLACKWELL, L., BOWNESS, P., BRENT, A., BROWN, A., BROXHOLME, J., BUCK, D., BURNHAM, K. L., BYRNE, H., CAMARA, S., CANDIDO FERREIRA, I., CHARLES, P., CHEN, W., CHEN, Y.-L., CHONG, A., CLUTTERBUCK, E. A., COLES, M., CONLON, C. P., CORNALL, R., CRIBBS, A. P., CURION, F., DAVENPORT, E. E., DAVIDSON, N., DAVIS, S., DENDROU, C. A., DEQUAIRE, J., DIB, L., DOCKER, J., DOLD, C., DONG, T., DOWNES, D., DRAKESMITH, H., DUNACHIE, S. J., DUNCAN, D. A., EIJSBOUTS, C., ESNOUF, R., ESPINOSA, A., ETHERINGTON, R., FAIRFAX, B., FAIRHEAD, R., FANG, H., FASSIH, S., FELLE, S., FERNANDEZ MENDOZA, M., FERREIRA, R., FISCHER, R., FOORD, T., FORROW, A., FRATER, J., FRIES, A., GALLARDO SANCHEZ, V., GARNER, L. C., GEEVES, C., GEORGIOU, D., GODFREY, L., GOLUBCHIK, T., GOMEZ VAZQUEZ, M., GREEN, A., HARPER, H., HARRINGTON, H. A., HEILIG, R., HESTER, S., HILL, J., HINDS, C., HIRD, C., HO, L.-P., HOEKZEMA, R., HOLLIS, B., HUGHES, J., HUTTON, P., JACKSON-WOOD, M. A., JAINARAYANAN, A., JAMES-BOTT, A., JANSEN, K., JEFFERY, K., JONES, E., JOSTINS, L., KERR, G., KIM, D., KLENERMAN, P., KNIGHT, J. C., KUMAR, V., et al. 2022. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell,* 185**,** 916-938.e58.

ALAMYAR, E., DUROUX, P., LEFRANC, M. P. & GIUDICELLI, V. 2012. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol,* 882**,** 569-604.

ALCOVER, A., ALARCON, B. & DI BARTOLO, V. 2018. Cell Biology of T Cell Receptor Expression and Regulation. *Annu Rev Immunol,* 36**,** 103-125.

ALISEYCHIK, M., PATRIKEEV, A., GUSEV, F., GRIGORENKO, A., ANDREEVA, T., BIRAGYN, A. & ROGAEV, E. 2020. Dissection of the Human T-Cell Receptor gamma Gene Repertoire in the Brain and Peripheral Blood Identifies Age- and Alzheimer's Disease-Associated Clonotype Profiles. *Front Immunol,* 11**,** 12.

ALLISON, T. J. & GARBOCZI, D. N. 2002. Structure of γδ T cell receptors and their recognition of non-peptide antigens. *Molecular Immunology,* 38**,** 1051-1061.

AMMER-HERRMENAU, C., KULKARNI, U., ANDREAS, N., UNGELENK, M., RAVENS, S., HUBNER, C., KATHER, A., KURTH, I., BAUER, M. & KAMRADT, T. 2019. Sepsis induces long-lasting impairments in CD4+ T-cell responses despite rapid numerical recovery of T-lymphocyte populations. *PLoS One,* 14**,** e0211716.

ANDREANI, T., SLOT, L. M., GABILLARD, S., STRUEBING, C., REIMERTZ, C., VEERANAGOUDA, Y., BAKKER, A. M., OLFATI-SABER, R., TOES, R. E. M., SCHERER, H. U., AUGÉ, F. & ŠIMAITĖ, D. 2022. Benchmarking computational methods for B-cell receptor reconstruction from single-cell RNA-seq data. *bioRxiv***,** 2022.03.24.485600.

ANDREU-BALLESTER, J. C., TORMO-CALANDIN, C., GARCIA-BALLESTEROS, C., PEREZ-GRIERA, J., AMIGO, V., ALMELA-QUILIS, A., RUIZ DEL CASTILLO, J., PENARROJA-OTERO, C. & BALLESTER, F. 2013. Association of gammadelta T cells with disease severity and mortality in septic patients. *Clin Vaccine Immunol,* 20**,** 738-46.

ANGELE, M. K., PRATSCHKE, S., HUBBARD, W. J. & CHAUDRY, I. H. 2014. Gender differences in sepsis: cardiovascular and immunological aspects. *Virulence,* 5**,** 12-9.

ANGUS, D. C., LINDE-ZWIRBLE, W. T., LIDICKER, J., CLERMONT, G., CARCILLO, J. & PINSKY, M. R. 2001. Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Critical Care Medicine,* 29**,** 1303-1310.

ANTCLIFFE, D. B., BURNHAM, K. L., AL-BEIDH, F., SANTHAKUMARAN, S., BRETT, S. J., HINDS, C. J., ASHBY, D., KNIGHT, J. C. & GORDON, A. C. 2019. Transcriptomic Signatures in Sepsis and a Differential Response to Steroids. From the VANISH Randomized Trial. *Am J Respir Crit Care Med,* 199**,** 980-986.

ANTONACCI, R., MASSARI, S., LINGUITI, G., CAPUTI JAMBRENGHI, A., GIANNICO, F., LEFRANC, M. P. & CICCARESE, S. 2020. Evolution of the T-Cell Receptor (TR) Loci in the Adaptive Immune Response: The Tale of the TRG Locus in Mammals. *Genes (Basel),* 11.

ARPIN, C., DE BOUTEILLER, O., RAZANAJAONA, D., FUGIER-VIVIER, I., BRIÈRE, F., BANCHEREAU, J., LEBECQUE, S. & LIU, Y. J. 1998. The normal counterpart of IgD myeloma cells in germinal center displays extensively mutated IgVH gene, Cmu-Cdelta switch, and lambda light chain expression. *J Exp Med,* 187**,** 1169-78.

ARVEY, A., ROWE, M., LEGUTKI, J. B., AN, G., GOLLAPUDI, A., LEI, A., COLSTON, B., PUTTERMAN, C., SMITH, D., STILES, J., TARASOW, T. & RAMAMOORTHY, P. 2020. Age-associated changes in the circulating human antibody repertoire are upregulated in autoimmunity. *Immun Ageing,* 17**,** 28.

ASGARI, S., MCLAREN, P. J., PEAKE, J., WONG, M., WONG, R., BARTHA, I., FRANCIS, J. R., ABARCA, K., GELDERMAN, K. A., AGYEMAN, P., AEBI, C., BERGER, C., FELLAY, J., SCHLAPBACH, L. J. & SWISS PEDIATRIC SEPSIS, S. 2016. Exome Sequencing Reveals Primary Immunodeficiencies in Children with Community-Acquired Pseudomonas aeruginosa Sepsis. *Front Immunol,* 7**,** 357.

AZEEZ, T., DHANESHWAR, K., ROMAN, S., SUDHAKAR, S., ABHISHEK, D. & SHARVAN, S. 2022. Glucocorticoid signaling mediates CD8+ T cell memory differentiation. *bioRxiv***,** 2022.07.19.500581.

BACKHAUS, O. 2018. Generation of Antibody Diversity.

BAGAEV, D. V., ZVYAGIN, I. V., PUTINTSEVA, E. V., IZRAELSON, M., BRITANOVA, O. V., CHUDAKOV, D. M. & SHUGAY, M. 2016. VDJviz: a versatile browser for immunogenomics data. *BMC Genomics,* 17**,** 453.

BAI, Y., WANG, D., LI, W., HUANG, Y., YE, X., WAITE, J., BARRY, T., EDELMANN, K. H., LEVENKOVA, N., GUO, C., SKOKOS, D., WEI, Y., MACDONALD, L. E. & FURY, W. 2018. Evaluation of the capacities of mouse TCR profiling from short read RNA-seq data. *PLoS One,* 13**,** e0207020.

BAKDASH, J. Z. & MARUSICH, L. R. 2017. Repeated Measures Correlation. *Front Psychol,* 8**,** 456.

BANDO, S. Y., BERTONHA, F. B., VIEIRA, S. E., DE OLIVEIRA, D. B. L., CHALUP, V. N., DURIGON, E. L., PALMEIRA, P., CURI, A. C. P., FARIA, C. S., ANTONANGELO, L., LAUTERBACH, G. D. P., REGALIO, F. A., CESAR, R. M., JR. & MOREIRA-FILHO, C. A. 2023. Blood leukocyte transcriptional modules and differentially expressed genes associated with disease severity and age in COVID-19 patients. *Sci Rep,* 13**,** 898.

BARKER, K. A., ETESAMI, N. S., SHENOY, A. T., ARAFA, E. I., LYON DE ANA, C., SMITH, N. M., MARTIN, I. M., GOLTRY, W. N., BARRON, A. M., BROWNING, J. L., KATHURIA, H., BELKINA, A. C., GUILLON, A., ZHONG, X., CROSSLAND, N. A., JONES, M. R., QUINTON, L. J. & MIZGERD, J. P. 2021. Lung-resident memory B cells protect against bacterial pneumonia. *J Clin Invest,* 131.

BARNAS, J. L., MEFFRE, E. & ANOLIK, J. H. 2021. B-cell biology, tolerance, and autoantibodies. *Lahita's Systemic Lupus Erythematosus.*

BARNATO, A. E., ALEXANDER, S. L., LINDE-ZWIRBLE, W. T. & ANGUS, D. C. 2008. Racial variation in the incidence, care, and outcomes of severe sepsis: analysis of population, patient, and hospital characteristics. *Am J Respir Crit Care Med,* 177**,** 279-84.

BASHFORD-ROGERS, R. J., NICOLAOU, K. A., BARTRAM, J., GOULDEN, N. J., LOIZOU, L., KOUMAS, L., CHI, J., HUBANK, M., KELLAM, P., COSTEAS, P. A. & VASSILIOU, G. S. 2016. Eye on the B-ALL: B-cell receptor repertoires reveal persistence of numerous B-lymphoblastic leukemia subclones from diagnosis to relapse. *Leukemia,* 30**,** 2312-2321.

BASHFORD-ROGERS, R. J., PALSER, A. L., HUNTLY, B. J., RANCE, R., VASSILIOU, G. S., FOLLOWS, G. A. & KELLAM, P. 2013. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Res,* 23**,** 1874-84.

BASHFORD-ROGERS, R. J. M., BERGAMASCHI, L., MCKINNEY, E. F., POMBAL, D. C., MESCIA, F., LEE, J. C., THOMAS, D. C., FLINT, S. M., KELLAM, P., JAYNE, D. R. W., LYONS, P. A. & SMITH, K. G. C. 2019. Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature,* 574**,** 122-126.

BASHFORD-ROGERS, R. J. M., PALSER, A. L., IDRIS, S. F., CARTER, L., EPSTEIN, M., CALLARD, R. E., DOUEK, D. C., VASSILIOU, G. S., FOLLOWS, G. A., HUBANK, M. & KELLAM, P. 2014. Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. *BMC Immunology,* 15**,** 29.

BATES, D., MÄCHLER, M., BOLKER, B. & WALKER, S. 2015. Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software,* 67.

BAUM, P. D., YOUNG, J. J., SCHMIDT, D., ZHANG, Q., HOH, R., BUSCH, M., MARTIN, J., DEEKS, S. & MCCUNE, J. M. 2012. Blood T-cell receptor diversity decreases during the course of HIV infection, but the potential for a diverse repertoire persists. *Blood,* 119**,** 3469-3477.

BAUMANN, S., HESS, J., EICHHORST, S. T., KRUEGER, A., ANGEL, P., KRAMMER, P. H. & KIRCHHOFF, S. 2003. An unexpected role for FosB in activation-induced cell death of T cells. *Oncogene,* 22**,** 1333-1339.

BERGKVIST, K. S., NØRGAARD, M. A., BØGSTED, M., SCHMITZ, A., NYEGAARD, M., GAIHEDE, M., BÆCH, J., GRØNHOLDT, M.-L., JENSEN, F. S., JOHANSEN, P., URUP, T., EL-GALALY, T. C., MADSEN, J., BØDKER, J. S., DYBKÆR, K. & JOHNSEN, H. E. 2016. Characterization of memory B cells from thymus and its impact for DLBCL classification. *Experimental Hematology,* 44**,** 982-990.e11.

BERMEJO-MARTIN, J. F. & GIAMARELLOS-BOURBOULIS, E. J. 2015. Endogenous immunoglobulins and sepsis: New perspectives for guiding replacement therapies. *International Journal of Antimicrobial Agents,* 46**,** S25-S28.

BERMEJO-MARTIN, J. F., RODRIGUEZ-FERNANDEZ, A., HERRAN-MONGE, R., ANDALUZ-OJEDA, D., MURIEL-BOMBIN, A., MERINO, P., GARCIA-GARCIA, M. M., CITORES, R., GANDIA, F., ALMANSA, R., BLANCO, J. & GROUP, G. 2014. Immunoglobulins IgG1, IgM and IgA: a synergistic team influencing survival in sepsis. *J Intern Med,* 276**,** 404-12.

BERNARD, B., RAJAMANICKAM, V., DUBAY, C., PIENING, B., ALONSO, E., JUTRIC, Z., TANG, E., NEWELL, P., HANSEN, P., MEDLER, T., GUNDERSON, A., YOUNG, K., BIFULCO, C., PUCLIOWSKA, J., CRITTENDEN, M. R. & GOUGH, M. J. 2020. Transcriptional and immunohistological assessment of immune infiltration in pancreatic cancer. *PLoS One,* 15**,** e0238380.

BINDER, C. J., HÖRKKÖ, S., DEWAN, A., CHANG, M.-K., KIEU, E. P., GOODYEAR, C. S., SHAW, P. X., PALINSKI, W., WITZTUM, J. L. & SILVERMAN, G. J. 2003. Pneumococcal vaccination decreases atherosclerotic lesion formation: molecular mimicry between Streptococcus pneumoniae and oxidized LDL. *Nature Medicine,* 9**,** 736-743.

BIRAM, A. & SHULMAN, Z. 2020. T cell help to B cells: Cognate and atypical interactions in peripheral and intestinal lymphoid tissues. *Immunol Rev,* 296**,** 36-47.

BISCHOF, J. & IBRAHIM, S. M. 2016. bcRep: R Package for Comprehensive Analysis of B Cell Receptor Repertoire Data. *PLOS ONE,* 11**,** e0161569.

BLACHLY, J. S., RUPPERT, A. S., ZHAO, W., LONG, S., FLYNN, J., FLINN, I., JONES, J., MADDOCKS, K., ANDRITSOS, L., GHIA, E. M., RASSENTI, L. Z., KIPPS, T. J., DE LA CHAPELLE, A. & BYRD, J. C. 2015. Immunoglobulin transcript sequence and somatic hypermutation computation from unselected RNA-seq reads in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A,* 112**,** 4322-7.

BLANDINO, R. & BAUMGARTH, N. 2019. Secreted IgM: New tricks for an old molecule. *J Leukoc Biol,* 106**,** 1021-1034.

BLUMENREICH, M. S. 1990. The White Blood Cell and Differential Count. 3rd ed.: Butterworths, Boston.

BOATENG, E. Y. & ABAYE, D. A. 2019. A Review of the Logistic Regression Model with Emphasis on Medical Research. *Journal of Data Analysis and Information Processing,* 07**,** 190-207.

BOEGEL, S., LÖWER, M., SCHÄFER, M., BUKUR, T., DE GRAAF, J., BOISGUÉRIN, V., TÜRECI, Ö., DIKEN, M., CASTLE, J. C. & SAHIN, U. 2012. HLA typing from RNA-Seq sequence reads. *Genome Medicine,* 4**,** 102.

BOES, M., PRODEUS, A. P., SCHMIDT, T., CARROLL, M. C. & CHEN, J. 1998. A critical role of natural immunoglobulin M in immediate defense against systemic bacterial infection. *Journal of Experimental Medicine,* 188**,** 2381-2386.

BOISMENU, R. & HAVRAN, W. L. 1994. Modulation of Epithelial Cell Growth by Intraepithelial γδ T Cells. *Science,* 266**,** 1253-1255.

BOLOTIN, D. A., POSLAVSKY, S., DAVYDOV, A. N., FRENKEL, F. E., FANCHI, L., ZOLOTAREVA, O. I., HEMMERS, S., PUTINTSEVA, E. V., OBRAZTSOVA, A. S., SHUGAY, M., ATAULLAKHANOV, R. I., RUDENSKY, A. Y., SCHUMACHER, T. N. & CHUDAKOV, D. M. 2017. Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol,* 35**,** 908-911.

BOLOTIN, D. A., POSLAVSKY, S., MITROPHANOV, I., SHUGAY, M., MAMEDOV, I. Z., PUTINTSEVA, E. V. & CHUDAKOV, D. M. 2015. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods,* 12**,** 380-1.

BOMANS, K., SCHENZ, J., SZTWIERTNIA, I., SCHAACK, D., WEIGAND, M. A. & UHLE, F. 2018. Sepsis Induces a Long-Lasting State of Trained Immunity in Bone Marrow Monocytes. *Front Immunol,* 9**,** 2685.

BONDT, A., HOEK, M., TAMARA, S., DE GRAAF, B., PENG, W., SCHULTE, D., VAN RIJSWIJCK, D. M. H., DEN BOER, M. A., GREISCH, J. F., VARKILA, M. R. J., SNIJDER, J., CREMER, O. L., BONTEN, M. J. M. & HECK, A. J. R. 2021. Human plasma IgG1 repertoires are simple, unique, and dynamic. *Cell Syst,* 12**,** 1131-1143 e5.

BONE, R. C., BALK, R. A., CERRA, F. B., DELLINGER, R. P., FEIN, A. M., KNAUS, W. A., SCHEIN, R. M. & SIBBALD, W. J. 1992. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest,* 101**,** 1644-55.

BOOMER, J. S., GREEN, J. M. & HOTCHKISS, R. S. 2014. The changing immune system in sepsis: is individualized immuno-modulatory therapy the answer? *Virulence,* 5**,** 45-56.

BOOMER, J. S., SHUHERK-SHAFFER, J., HOTCHKISS, R. S. & GREEN, J. M. 2012. A prospective analysis of lymphocyte phenotype and function over the course of acute sepsis. *Crit Care,* 16**,** R112.

BOURNAZOS, S. & RAVETCH, J. V. 2017. Diversification of IgG effector functions. *Int Immunol,* 29**,** 303-310.

BRADY, J., HORIE, S. & LAFFEY, J. G. 2020. Role of the adaptive immune response in sepsis. *Intensive Care Med Exp,* 8**,** 20.

BRAY, A., KAMPOURAKI, E., WINTER, A., JESUTHASAN, A., MESSER, B. & GRAZIADIO, S. 2020. High Variability in Sepsis Guidelines in UK: Why Does It Matter? *Int J Environ Res Public Health,* 17.

BREBNER, J. A. & STOCKLEY, R. A. 2013. Polyclonal free light chains: a biomarker of inflammatory disease or treatment target? *F1000 Med Rep,* 5**,** 4.

BRITANOVA, O. V., PUTINTSEVA, E. V., SHUGAY, M., MERZLYAK, E. M., TURCHANINOVA, M. A., STAROVEROV, D. B., BOLOTIN, D. A., LUKYANOV, S., BOGDANOVA, E. A., MAMEDOV, I. Z., LEBEDEV, Y. B. & CHUDAKOV, D. M. 2014. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol,* 192**,** 2689-98.

BRUNI, E., CIMINO, M. M., DONADON, M., CARRIERO, R., TERZOLI, S., PIAZZA, R., RAVENS, S., PRINZ, I., CAZZETTA, V., MARZANO, P., KUNDERFRANCO, P., PEANO, C., SOLDANI, C., FRANCESCHINI, B., COLOMBO, F. S., GARLANDA, C., MANTOVANI, A., TORZILLI, G., MIKULAK, J. & MAVILIO, D. 2022. Intrahepatic CD69⁺Vδ1 T cells re-circulate in the blood of patients with metastatic colorectal cancer and limit tumor progression. *Journal for ImmunoTherapy of Cancer,* 10**,** e004579.

BUCHHOLZ, V. R., SCHUMACHER, T. N. & BUSCH, D. H. 2016. T Cell Fate at the Single-Cell Level. *Annu Rev Immunol,* 34**,** 65-92.

BUDEUS, B., KIBLER, A., BRAUSER, M., HOMP, E., BRONISCHEWSKI, K., ROSS, J. A., GÖRGENS, A., WENIGER, M. A., DUNST, J., KRESLAVSKY, T., CASTRO, S. V. D. C., MURKE, F., OAKES, C. C., RUSCH, P., ANDRIKOS, D., KERN, P., KÖNINGER, A., LINDEMANN, M., JOHANSSON, P., HANSEN, W., LUNDELL, A.-C., RUDIN, A., DÜRIG, J., GIEBEL, B., HOFFMANN, D., KÜPPERS, R. & SEIFERT, M. 2020. Human neonatal B cell immunity differs from the adult version by conserved Ig repertoires and rapid, but transient response dynamics. *bioRxiv***,** 2020.08.11.245985.

BURBELO, P. D., SEAM, N., GROOT, S., CHING, K. H., HAN, B. L., MEDURI, G. U., IADAROLA, M. J. & SUFFREDINI, A. F. 2010. Rapid induction of autoantibodies during ARDS and septic shock. *J Transl Med,* 8**,** 97.

BURNHAM, K. L., DAVENPORT, E. E., RADHAKRISHNAN, J., HUMBURG, P., GORDON, A. C., HUTTON, P., SVOREN-JABALERA, E., GARRARD, C., HILL, A. V. S., HINDS, C. J. & KNIGHT, J. C. 2017. Shared and Distinct Aspects of the Sepsis Transcriptomic Response to Fecal Peritonitis and Pneumonia. *American Journal of Respiratory and Critical Care Medicine,* 196**,** 328-339.

BURNHAM, K. P. & ANDERSON, D. R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*, Springer Verlag.

CAIRO, C., ARMSTRONG, C. L., CUMMINGS, J. S., DEETZ, C. O., TAN, M., LU, C., DAVIS, C. E. & PAUZA, C. D. 2010. Impact of age, gender, and race on circulating γδ T cells. *Hum Immunol,* 71**,** 968-75.

CALLAN, M. F. C., STEVEN, N., KRAUSA, P., WILSON, J. D. K., MOSS, P. A. H., GILLESPIE, G. M., BELL, J. I., RICKINSON, A. B. & MCMICHAEL, A. J. 1996. Large clonal expansions of CD8+ T cells in acute infectious mononucleosis. *Nature Medicine,* 2**,** 906-911.

CANO-GAMEZ, E., BURNHAM, K. L., GOH, C., ALLCOCK, A., MALICK, Z. H., OVEREND, L., KWOK, A., SMITH, D. A., PETERS-SENGERS, H., ANTCLIFFE, D., MCKECHNIE, S., SCICLUNA, B. P., VAN DER POLL, T., GORDON, A. C., HINDS, C. J., DAVENPORT, E. E., KNIGHT, J. C., WEBSTER, N., GALLEY, H., TAYLOR, J., HALL, S., ADDISON, J., ROUGHTON, S., TENNANT, H., GULERI, A., WADDINGTON, N., ARAWWAWALA, D., DURCAN, J., SHORT, A., SWAN, K., WILLIAMS, S., SMOLEN, S., MITCHELL-INWANG, C., GORDON, T., ERRINGTON, E., TEMPLETON, M., VENATESH, P., WARD, G., MCCAULEY, M., BAUDOUIN, S., HIGHAM, C., SOAR, J., GRIER, S., HALL, E., BRETT, S., KITSON, D., WILSON, R., MOUNTFORD, L., MORENO, J., HALL, P., HEWLETT, J., MCKECHNIE, S., GARRARD, C., MILLO, J., YOUNG, D., HUTTON, P., PARSONS, P., SMITHS, A., FARAS-ARRAYA, R., SOAR, J., RAYMODE, P., THOMPSON, J., BOWREY, S., KAZEMBE, S., RICH, N., ANDREOU, P., HALES, D., ROBERTS, E., FLETCHER, S., ROSBERGEN, M., GLISTER, G., CUESTA, J. M., BION, J., MILLAR, J., PERRY, E. J., WILLIS, H., MITCHELL, N., RUEL, S., CARRERA, R., WILDE, J., NILSON, A., LEES, S., KAPILA, A., JACQUES, N., ATKINSON, J., BROWN, A., PROWSE, H., KRIGE, A., BLAND, M., BULLOCK, L., HARRISON, D., MILLS, G., HUMPHREYS, J., ARMITAGE, K., LAHA, S., BALDWIN, J., WALSH, A., DOHERTY, N., DRAGE, S., ORTIZ-RUIZ DE GORDOA, L., et al. 2022. An immune dysfunction score for stratification of patients with acute infection based on whole-blood gene expression. *Science Translational Medicine,* 14**,** eabq4433.

CANTINI, L., ZAKERI, P., HERNANDEZ, C., NALDI, A., THIEFFRY, D., REMY, E. & BAUDOT, A. 2021. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature Communications,* 12**,** 124.

CAO, C., YU, M. & CHAI, Y. 2019. Pathological alteration and therapeutic implications of sepsis-induced immune cell apoptosis. *Cell Death Dis,* 10**,** 782.

CARTER, R. H. 2006. B Cells in Health and Disease. *Mayo Clinic Proceedings,* 81**,** 377-384.

CASTRO, C. D. & FLAJNIK, M. F. 2014. Putting J chain back on the map: how might its expression define plasma cell development? *J Immunol,* 193**,** 3248-55.

CECCONI, M., EVANS, L., LEVY, M. & RHODES, A. 2018. Sepsis and septic shock. *The Lancet,* 392**,** 75-87.

CERUTTI, A. 2008. The regulation of IgA class switching. *Nat Rev Immunol,* 8**,** 421-34.

CHAI, K., ZHANG, X., TANG, H., GU, H., YE, W., WANG, G., CHEN, S., WAN, F., LIANG, J. & SHEN, D. 2022. The Application of Consensus Weighted Gene Co-expression Network Analysis to Comparative Transcriptome Meta-Datasets of Multiple Sclerosis in Gray and White Matter. *Frontiers in Neurology,* 13.

CHAKRABORTY, J., PALIT, K. & DAS, S. 2022. Chapter 5 - Metagenomic approaches to study the culture-independent bacterial diversity of a polluted environment—a case study on north-eastern coast of Bay of Bengal, India. *In:* DAS, S. & DASH, H. R. (eds.) *Microbial Biodegradation and Bioremediation (Second Edition).* Elsevier.

CHARLSON, M. E., POMPEI, P., ALES, K. L. & MACKENZIE, C. R. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases,* 40**,** 373-383.

CHARRAD, M., GHAZZALI, N., BOITEAU, V. & NIKNAFS, A. 2012. NbClust Package. An examination of indices for determining the number of clusters.

CHAUDHARY, N. & WESEMANN, D. R. 2018. Analyzing Immunoglobulin Repertoires. *Front Immunol,* 9**,** 462.

CHEN, C., NAGY, Z., LUNING PRAK, E. & WEIGERT, M. 1995. Immunoglobulin heavy chain gene replacement: A mechanism of receptor editing. *Immunity,* 3**,** 747-755.

CHEN, G., YANG, X., KO, A., SUN, X., GAO, M., ZHANG, Y., SHI, A., MARIUZZA, R. A. & WENG, N.-P. 2017. Sequence and Structural Analyses Reveal Distinct and Highly Diverse Human CD8+ TCR Repertoires to Immunodominant Viral Antigens. *Cell Reports,* 19**,** 569-583.

CHENG, M. & HU, S. 2017. Lung-resident γδ T cells and their roles in lung diseases. *Immunology,* 151**,** 375-384.

CHENG, P.-L., CHEN, H.-H., JIANG, Y.-H., HSIAO, T.-H., WANG, C.-Y., WU, C.-L., KO, T.-M. & CHAO, W.-C. 2021. Using RNA-Seq to Investigate Immune-Metabolism Features in Immunocompromised Patients With Sepsis. *Frontiers in Medicine,* 8.

CHENG, P., LIU, T., ZHOU, W.-Y., ZHUANG, Y., PENG, L.-S., ZHANG, J.-Y., YIN, Z.-N., MAO, X.-H., GUO, G., SHI, Y. & ZOU, Q.-M. 2012. Role of gamma-delta T cells in host response against Staphylococcus aureus-induced pneumonia. *BMC Immunology,* 13**,** 38.

CHI, X., LI, Y. & QIU, X. 2020. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology,* 160**,** 233-247.

CHIU, C. Y. & MILLER, S. A. 2019. Clinical metagenomics. *Nat Rev Genet,* 20**,** 341-355.

CHRISTIE, S. M., FIJEN, C. & ROTHENBERG, E. 2022. V(D)J Recombination: Recent Insights in Formation of the Recombinase Complex and Recruitment of DNA Repair Machinery. *Front Cell Dev Biol,* 10**,** 886718.

CHRONISTER, W. D., CRINKLAW, A., MAHAJAN, S., VITA, R., KOŞALOĞLU-YALÇıN, Z., YAN, Z., GREENBAUM, J. A., JESSEN, L. E., NIELSEN, M., CHRISTLEY, S., COWELL, L. G., SETTE, A. & PETERS, B. 2021. TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors. *Frontiers in Immunology,* 12.

CHUN, K., SYNDERGAARD, C., DAMAS, C., TRUBEY, R., MUKINDARAJ, A., QIAN, S., JIN, X., BRESLOW, S. & NIEMZ, A. 2015. Sepsis Pathogen Identification. *J Lab Autom,* 20**,** 539-61.

CIUPE, S. M., DEVLIN, B. H., MARKERT, M. L. & KEPLER, T. B. 2013. Quantification of total T-cell receptor diversity by flow cytometry and spectratyping. *BMC Immunology,* 14**,** 35.

CLARK, B. L. & THOMAS, P. G. 2020. A Cell for the Ages: Human γδ T Cells across the Lifespan. *Int J Mol Sci,* 21.

CLEMENTS, T. W., TOLONEN, M., BALL, C. G. & KIRKPATRICK, A. W. 2021. Secondary Peritonitis and Intra-Abdominal Sepsis: An Increasingly Global Disease in Search of Better Systemic Therapies. *Scandinavian Journal of Surgery,* 110**,** 139-149.

COLLINS, A. M. & WATSON, C. T. 2018. Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire. *Front Immunol,* 9**,** 2249.

COMEAU, K., PARADIS, P. & SCHIFFRIN, E. L. 2020. Human and murine memory gammadelta T cells: Evidence for acquired immune memory in bacterial and viral infections and autoimmunity. *Cell Immunol,* 357**,** 104217.

CORTINA-CEBALLOS, B., GODOY-LOZANO, E. E., TÉLLEZ-SOSA, J., OVILLA-MUÑOZ, M., SÁMANO-SÁNCHEZ, H., AGUILAR-SALGADO, A., GÓMEZ-BARRETO, R. E., VALDOVINOS-TORRES, H., LÓPEZ-MARTÍNEZ, I., APARICIO-ANTONIO, R., RODRÍGUEZ, M. H. & MARTÍNEZ-BARNETCHE, J. 2015. Longitudinal analysis of the peripheral B cell repertoire reveals unique effects of immunization with a new influenza virus strain. *Genome Medicine,* 7**,** 124.

COSTELLO, M., FLEHARTY, M., ABREU, J., FARJOUN, Y., FERRIERA, S., HOLMES, L., GRANGER, B., GREEN, L., HOWD, T., MASON, T., VICENTE, G., DASILVA, M., BRODEUR, W., DESMET, T., DODGE, S., LENNON, N. J. & GABRIEL, S. 2018. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics,* 19**,** 332.

CSARDI, G. & NEPUSZ, T. 2005. The Igraph Software Package for Complex Network Research. *InterJournal,* Complex Systems**,** 1695.

CYSTER, J. G. & ALLEN, C. D. C. 2019. B Cell Responses: Cell Interaction Dynamics and Decisions. *Cell,* 177**,** 524-540.

DARLOW, J. M. & STOTT, D. I. 2005. V(H) replacement in rearranged immunoglobulin genes. *Immunology,* 114**,** 155-65.

DAVENPORT, E. E., BURNHAM, K. L., RADHAKRISHNAN, J., HUMBURG, P., HUTTON, P., MILLS, T. C., RAUTANEN, A., GORDON, A. C., GARRARD, C., HILL, A. V. S., HINDS, C. J. & KNIGHT, J. C. 2016. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *The Lancet Respiratory Medicine,* 4**,** 259-271.

DE BOURCY, C. F. A., DEKKER, C. L., DAVIS, M. M., NICOLLS, M. R. & QUAKE, S. R. 2017. Dynamics of the human antibody repertoire after B cell depletion in systemic sclerosis. *Science Immunology,* 2**,** eaan8289.

DE GREEF, P. C., OAKES, T., GERRITSEN, B., ISMAIL, M., HEATHER, J. M., HERMSEN, R., CHAIN, B. & DE BOER, R. J. 2020. The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes. *Elife,* 9.

DE LA ROSA, K., TAN, J., PICCOLI, L., FOGLIERINI PEREZ, M., BARBIERI, S., CHEN, Y., SILACCI-FREGNI, C., WOLF, T., JARROSSAY, D., ANDERLE, M., ABDI, A., NDUNGU, F., DOUMBO, O., TRAORE, B., TRAN, T., JONGO, S., ZENKLUSEN, I., CROMPTON, P., DAUBENBERGER, C. & LANZAVECCHIA, A. 2017. Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. *Nature,* 548.

DE LOUCHE, C. D. & ROGHANIAN, A. 2022. Human inhibitory leukocyte Ig-like receptors: from immunotolerance to immunotherapy. *JCI Insight,* 7.

DE MOL, J., KUIPER, J., TSIANTOULAS, D. & FOKS, A. C. 2021. The Dynamics of B Cell Aging in Health and Disease. *Front Immunol,* 12**,** 733566.

DE SOUZA COSTA, M. F., BASTOS TRIGO DE NEGREIROS, C., UGARTE BORNSTEIN, V., HEMMI VALENTE, R., MENGEL, J., HENRIQUES, M. D. G., FARIAS BENJAMIM, C. & PENIDO, C. 2015. Murine IL-17+ Vγ4 T lymphocytes accumulate in the lungs and play a protective role during severe sepsis. *BMC Immunology,* 16**,** 36.

DEKOSKY, B. J., IPPOLITO, G. C., DESCHNER, R. P., LAVINDER, J. J., WINE, Y., RAWLINGS, B. M., VARADARAJAN, N., GIESECKE, C., DÖRNER, T., ANDREWS, S. F., WILSON, P. C., HUNICKE-SMITH, S. P., WILLSON, C. G., ELLINGTON, A. D. & GEORGIOU, G. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnology,* 31**,** 166-169.

DEKOSKY, B. J., KOJIMA, T., RODIN, A., CHARAB, W., IPPOLITO, G. C., ELLINGTON, A. D. & GEORGIOU, G. 2015. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nature Medicine,* 21**,** 86-91.

DELANO, M. J. & WARD, P. A. 2016. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol Rev,* 274**,** 330-353.

DENSTAEDT, S. J., SINGER, B. H. & STANDIFORD, T. J. 2018. Sepsis and Nosocomial Infection: Patient Characteristics, Mechanisms, and Modulation. *Front Immunol,* 9**,** 2446.

DESEKE, M. & PRINZ, I. 2020. Ligand recognition by the gammadelta TCR and discrimination between homeostasis and stress conditions. *Cell Mol Immunol,* 17**,** 914-924.

DEWITT, W. S., III, LINDAU, P., EMERSON, R. O., SHERWOOD, A., WILLIAMSON, D., RIEDER, M. J., CHUNG, M., CARLSON, C. S. & ROBINS, H. 2013. Assessing B Lymphocyte Clonal Diversity, Expansion, and Convergent Evolution By High-Throughput Sequencing Of Rearranged IGH Segments From Naïve and Memory Repertoires. *Blood,* 122**,** 1045-1045.

DI NOIA, J. M. & NEUBERGER, M. S. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem,* 76**,** 1-22.

DIMEGLIO, M., DUBENSKY, J., SCHADT, S., POTDAR, R. & LAUDANSKI, K. 2018. Factors Underlying Racial Disparities in Sepsis Management. *Healthcare (Basel),* 6.

DIOSA-TORO, M., VELILLA, P. A., RUGELES, M. T. & JAIMES, F. A. 2013. Increased frequency of NK and gamma delta 1 T cells in a cohort of patients with sepsis*. *Infectio,* 17**,** 177-184.

DISPENZIERI, A., KATZMANN, J. A., KYLE, R. A., LARSON, D. R., THERNEAU, T. M., COLBY, C. L., CLARK, R. J., MEAD, G. P., KUMAR, S., MELTON, L. J. & RAJKUMAR, S. V. 2012. Use of Nonclonal Serum Immunoglobulin Free Light Chains to Predict Overall Survival in the General Population. *Mayo Clinic Proceedings,* 87**,** 517-523.

DOLIN, H. H., PAPADIMOS, T. J., CHEN, X. & PAN, Z. K. 2019. Characterization of Pathogenic Sepsis Etiologies and Patient Profiles: A Novel Approach to Triage and Treatment. *Microbiol Insights,* 12**,** 1178636118825081.

DONG, ZHENG, L., LIN, J., ZHANG, B., ZHU, Y., LI, N., XIE, S., WANG, Y., GAO, N. & HUANG, Z. 2019. Structural basis of assembly of the human T cell receptor-CD3 complex. *Nature,* 573**,** 546-552.

DONG, C. 2021. Cytokine Regulation and Function in T Cells. *Annu Rev Immunol,* 39**,** 51-76.

DOPFER, E. P., HARTL, F. A., OBERG, H. H., SIEGERS, G. M., YOUSEFI, O. S., KOCK, S., FIALA, G. J., GARCILLAN, B., SANDSTROM, A., ALARCON, B., REGUEIRO, J. R., KABELITZ, D., ADAMS, E. J., MINGUET, S., WESCH, D., FISCH, P. & SCHAMEL, W. W. A. 2014. The CD3 conformational change in the gammadelta T cell receptor is not triggered by antigens but can be enforced to enhance tumor killing. *Cell Rep,* **7,** 1704-1715.

DUAN, L. W., QU, J. L., WAN, J., XU, Y. H., SHAN, Y., WU, L. X., ZHENG, J. H., JIANG, W. W., CHEN, Q. T., ZHU, Y., ZHOU, J., YU, W. B., PEI, L., SONG, X., LI, W. F. & LIN, Z. F. 2021. Effects of viral infection and microbial diversity on patients with sepsis: A retrospective study based on metagenomic next-generation sequencing. *World J Emerg Med,* 12**,** 29-35.

DUAN, S., JIAO, Y., WANG, J., TANG, D., XU, S., WANG, R., JIANG, T., SHAO, J., HE, Z. & YU, W. 2020. Impaired B-Cell Maturation Contributes to Reduced B Cell Numbers and Poor Prognosis in Sepsis. *Shock,* 54**,** 70-77.

DUNN-WALTERS, D. K. 2016. The ageing human B cell repertoire: a failure of selection? *Clin Exp Immunol,* 183**,** 50-6.

DYLKE, J., LOPES, J., DANG-LAWSON, M., MACHTALER, S. & MATSUUCHI, L. 2007. Role of the extracellular and transmembrane domain of Ig-alpha/beta in assembly of the B cell antigen receptor (BCR). *Immunol Lett,* 112**,** 47-57.

ECKMANN, C. & BASSETTI, M. 2014. Prognostic factors for mortality in (fecal) peritonitis: back to the roots! *Intensive Care Medicine,* 40**,** 269-271.

EDWARDS, J. C. W. & CAMBRIDGE, G. 2006. B-cell targeting in rheumatoid arthritis and other autoimmune diseases. *Nature Reviews Immunology,* 6**,** 394-403.

EGOROV, E. S., KASATSKAYA, S. A., ZUBOV, V. N., IZRAELSON, M., NAKONECHNAYA, T. O., STAROVEROV, D. B., ANGIUS, A., CUCCA, F., MAMEDOV, I. Z., ROSATI, E., FRANKE, A., SHUGAY, M., POGORELYY, M. V., CHUDAKOV, D. M. & BRITANOVA, O. V. 2018. The Changing Landscape of Naive T Cell Receptor Repertoire With Human Aging. *Front Immunol,* 9**,** 1618.

EIBEL, H., KRAUS, H., SIC, H., KIENZLER, A. K. & RIZZI, M. 2014. B cell biology: an overview. *Curr Allergy Asthma Rep,* 14**,** 434.

EICHHORST, S. T., MÜLLER, M., LI-WEBER, M., SCHULZE-BERGKAMEN, H., ANGEL, P. & KRAMMER, P. H. 2000. A Novel AP-1 Element in the CD95 Ligand Promoter Is Required for Induction of Apoptosis in Hepatocellular Carcinoma Cells upon Treatment with Anticancer Drugs. *Molecular and Cellular Biology,* 20**,** 7826-7837.

EIL, R., VODNALA, S. K., CLEVER, D., KLEBANOFF, C. A., SUKUMAR, M., PAN, J. H., PALMER, D. C., GROS, A., YAMAMOTO, T. N., PATEL, S. J., GUITTARD, G. C., YU, Z., CARBONARO, V., OKKENHAUG, K., SCHRUMP, D. S., LINEHAN, W. M., ROYCHOUDHURI, R. & RESTIFO, N. P. 2016. Ionic immune suppression within the tumour microenvironment limits T cell effector function. *Nature,* 537**,** 539-543.

ENGELHART, S., GLYNN, R. J. & SCHUR, P. H. 2017. Disease associations with isolated elevations of each of the four IgG subclasses. *Semin Arthritis Rheum,* 47**,** 276-280.

ENGOREN, M., JEWELL, E. S., DOUVILLE, N., MOSER, S., MAILE, M. D. & BAUER, M. E. 2022. Genetic variants associated with sepsis. *PLoS One,* 17**,** e0265052.

EZHUTHACHAN, I., SHAPERO, M., FISHBEIN, J. & KAPLAN, B. 2021. Lower IgM levels after rituximab use is associated with sepsis. *Journal of Allergy and Clinical Immunology,* 147**,** AB73.

FALCAO-HOLANDA, R. B., BRUNIALTI, M. K. C., JASIULIONIS, M. G. & SALOMAO, R. 2021. Epigenetic Regulation in Sepsis, Role in Pathophysiology and Therapeutic Perspective. *Front Med (Lausanne),* 8**,** 685333.

FANG, H., KNEZEVIC, B., BURNHAM, K. L. & KNIGHT, J. C. 2016. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Medicine,* 8**,** 129.

FANG, H., YAMAGUCHI, R., LIU, X., DAIGO, Y., YEW, P. Y., TANIKAWA, C., MATSUDA, K., IMOTO, S., MIYANO, S. & NAKAMURA, Y. 2014. Quantitative T cell repertoire analysis by deep cDNA sequencing of T cell receptor alpha and beta chains using next-generation sequencing (NGS). *Oncoimmunology,* 3**,** e968467.

FANG, M., SU, Z., ABOLHASSANI, H., ZHANG, W., JIANG, C., CHENG, B., LUO, L., WU, J., WANG, S., LIN, L., WANG, X., WANG, L., AGHAMOHAMMADI, A., LI, T., ZHANG, X., HAMMARSTRÖM, L. & LIU, X. 2022. T Cell Repertoire Abnormality in Immunodeficiency Patients with DNA Repair and Methylation Defects. *Journal of Clinical Immunology,* 42**,** 375-393.

FANG, Q., WANG, Q., ZHOU, Z. & XIE, A. 2021. Consensus analysis via weighted gene co-expression network analysis (WGCNA) reveals genes participating in early phase of acute respiratory distress syndrome (ARDS) induced by sepsis. *Bioengineered,* 12**,** 1161-1172.

FARBER, D. L., YUDANIN, N. A. & RESTIFO, N. P. 2014. Human memory T cells: generation, compartmentalization and homeostasis. *Nat Rev Immunol,* 14**,** 24-35.

FARMANBAR, A., KNELLER, R. & FIROUZI, S. 2019. RNA sequencing identifies clonal structure of T-cell repertoires in patients with adult T-cell leukemia/lymphoma. *npj Genomic Medicine,* 4**,** 10.

FAROUNI, R., DJAMBAZIAN, H., FERRI, L. E., RAGOUSSIS, J. & NAJAFABADI, H. S. 2020. Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. *Nat Commun,* 11**,** 2704.

FERWERDA, B., ALONSO, S., BANAHAN, K., MCCALL, M. B., GIAMARELLOS-BOURBOULIS, E. J., RAMAKERS, B. P., MOUKTAROUDI, M., FAIN, P. R., IZAGIRRE, N., SYAFRUDDIN, D., CRISTEA, T., MOCKENHAUPT, F. P., TROYE-BLOMBERG, M., KUMPF, O., MAIGA, B., DOLO, A., DOUMBO, O., SUNDARESAN, S., BEDU-ADDO, G., VAN CREVEL, R., HAMANN, L., OH, D. Y., SCHUMANN, R. R., JOOSTEN, L. A., DE LA RÚA, C., SAUERWEIN, R., DRENTH, J. P., KULLBERG, B. J., VAN DER VEN, A. J., HILL, A. V., PICKKERS, P., VAN DER MEER, J. W., O'NEILL, L. A. & NETEA, M. G. 2009. Functional and genetic evidence that the Mal/TIRAP allele variant 180L has been selected by providing protection against septic shock. *Proc Natl Acad Sci U S A,* 106**,** 10272-7.

FICHTNER, A. S., RAVENS, S. & PRINZ, I. 2020. Human γδ TCR Repertoires in Health and Disease. *Cells,* 9.

FINFER, S., VENKATESH, B., HOTCHKISS, R. S. & SASSON, S. C. 2022. Lymphopenia in sepsis—an acquired immunodeficiency? *Immunology & Cell Biology,* n/a.

FISCHER, D. S., WU, Y., SCHUBERT, B. & THEIS, F. J. 2020. Predicting antigen specificity of single T cells based on TCR CDR3 regions. *Molecular Systems Biology,* 16**,** e9416.

FISHER, C. J., JR., AGOSTI, J. M., OPAL, S. M., LOWRY, S. F., BALK, R. A., SADOFF, J. C., ABRAHAM, E., SCHEIN, R. M. & BENJAMIN, E. 1996. Treatment of septic shock with the tumor necrosis factor receptor:Fc fusion protein. The Soluble TNF Receptor Sepsis Study Group. *N Engl J Med,* 334**,** 1697-702.

FLYNN MAKIC, M. B. 2019. Sepsis: Noninvasive Assessment. *J Perianesth Nurs,* 34**,** 861-863.

FORT, A., PANOUSIS, N. I., GARIERI, M., ANTONARAKIS, S. E., LAPPALAINEN, T., DERMITZAKIS, E. T. & DELANEAU, O. 2017. MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics,* 33**,** 1895-1897.

FOZZA, C., BARRAQUEDDU, F., CORDA, G., CONTINI, S., VIRDIS, P., DORE, F., BONFIGLI, S. & LONGINOTTI, M. 2017. Study of the T-cell receptor repertoire by CDR3 spectratyping. *J Immunol Methods,* 440**,** 1-11.

FRANCOIS, B., JEANNET, R., DAIX, T., WALTON, A. H., SHOTWELL, M. S., UNSINGER, J., MONNERET, G., RIMMELE, T., BLOOD, T., MORRE, M., GREGOIRE, A., MAYO, G. A., BLOOD, J., DURUM, S. K., SHERWOOD, E. R. & HOTCHKISS, R. S. 2018. Interleukin-7 restores lymphocytes in septic shock: the IRIS-7 randomized clinical trial. *JCI Insight,* 3.

FURST, D. E. 2009. Serum immunoglobulins and risk of infection: how low can you go? *Semin Arthritis Rheum,* 39**,** 18-29.

GALSON, J. D., CLUTTERBUCK, E. A., TRÜCK, J., RAMASAMY, M. N., MÜNZ, M., FOWLER, A., CERUNDOLO, V., POLLARD, A. J., LUNTER, G. & KELLY, D. F. 2015a. BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. *Immunology & Cell Biology,* 93**,** 885-895.

GALSON, J. D., SCHAETZLE, S., BASHFORD-ROGERS, R. J. M., RAYBOULD, M. I. J., KOVALTSUK, A., KILPATRICK, G. J., MINTER, R., FINCH, D. K., DIAS, J., JAMES, L. K., THOMAS, G., LEE, W.-Y. J., BETLEY, J., CAVLAN, O., LEECH, A., DEANE, C. M., SEOANE, J., CALDAS, C., PENNINGTON, D. J., PFEFFER, P. & OSBOURN, J. 2020. Deep Sequencing of B Cell Receptor Repertoires From COVID-19 Patients Reveals Strong Convergent Immune Signatures. *Frontiers in Immunology,* 11.

GALSON, J. D., TRÜCK, J., FOWLER, A., CLUTTERBUCK, E. A., MÜNZ, M., CERUNDOLO, V., REINHARD, C., VAN DER MOST, R., POLLARD, A. J., LUNTER, G. & KELLY, D. F. 2015b. Analysis of B Cell Repertoire Dynamics Following Hepatitis B Vaccination in Humans, and Enrichment of Vaccine-specific Antibody Sequences. *EBioMedicine,* 2**,** 2070-2079.

GARCILLÁN, B., MARIN, A. V., JIMÉNEZ-REINOSO, A., BRIONES, A. C., MUÑOZ-RUIZ, M., GARCÍA-LEÓN, M. J., GIL, J., ALLENDE, L. M., MARTÍNEZ-NAVES, E., TORIBIO, M. L. & REGUEIRO, J. R. 2015. γδ T Lymphocytes in the Diagnosis of Human T Cell Receptor Immunodeficiencies. *Front Immunol,* 6**,** 20.

GARETH JAMES, D. W. T. H. R. T. 2013. *An introduction to statistical learning : with applications in R*, New York : Springer, [2013] ©2013.

GAROFALO, S., GIOVAGNOLI, S., ORSONI, M., STARITA, F. & BENASSI, M. 2022. Interaction effect: Are you doing the right thing? *PLoS One,* 17**,** e0271668.

GATTO, D. & BRINK, R. 2010. The germinal center reaction. *J Allergy Clin Immunol,* 126**,** 898-907; quiz 908-9.

GERLACH, C., VAN HEIJST, J. W., SWART, E., SIE, D., ARMSTRONG, N., KERKHOVEN, R. M., ZEHN, D., BEVAN, M. J., SCHEPERS, K. & SCHUMACHER, T. N. 2010. One naive T cell, multiple fates in CD8+ T cell differentiation. *J Exp Med,* 207**,** 1235-46.

GHRAICHY, M., VON NIEDERHÄUSERN, V., KOVALTSUK, A., GALSON, J. D., DEANE, C. M. & TRÜCK, J. 2021. Different B cell subpopulations show distinct patterns in their IgH repertoire metrics. *eLife,* 10**,** e73111.

GIAMARELLOS-BOURBOULIS, E. J., APOSTOLIDOU, E., LADA, M., PERDIOS, I., GATSELIS, N. K., TSANGARIS, I., GEORGITSI, M., BRISTIANOU, M., KANNI, T., SERETI, K., KYPRIANOU, M. A., KOTANIDOU, A., ARMAGANIDIS, A. & HELLENIC SEPSIS STUDY, G. 2013. Kinetics of circulating immunoglobulin M in sepsis: relationship with final outcome. *Critical care (London, England),* 17**,** R247-R247.

GIBSON, K. L., WU, Y. C., BARNETT, Y., DUGGAN, O., VAUGHAN, R., KONDEATIS, E., NILSSON, B. O., WIKBY, A., KIPLING, D. & DUNN-WALTERS, D. K. 2009. B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell,* 8**,** 18-25.

GOH, C., BURNHAM, K. L., ANSARI, M. A., DE CESARE, M., GOLUBCHIK, T., HUTTON, P., OVEREND, L. E., DAVENPORT, E. E., HINDS, C. J., BOWDEN, R. & KNIGHT, J. C. 2020. Epstein-Barr virus reactivation in sepsis due to community-acquired pneumonia is associated with increased morbidity and an immunosuppressed host transcriptomic endotype. *Sci Rep,* 10**,** 9838.

GOLBECK, J. 2013. Chapter 3 - Network Structure and Measures. *In:* GOLBECK, J. (ed.) *Analyzing the Social Web.* Boston: Morgan Kaufmann.

GORONZY, J. J., ZETTL, A. & WEYAND, C. M. 1998. T cell receptor repertoire in rheumatoid arthritis. *Int Rev Immunol,* 17**,** 339-63.

GRASSEAU, A., BOUDIGOU, M., MICHEE-COSPOLITE, M., DELALOY, C., MIGNEN, O., JAMIN, C., CORNEC, D., PERS, J. O., LE POTTIER, L. & HILLION, S. 2021. The diversity of the plasmablast signature across species and experimental conditions: A meta-analysis. *Immunology,* 164**,** 120-134.

GRAY, J. I., WESTERHOF, L. M. & MACLEOD, M. K. L. 2018. The roles of resident, central and effector memory CD4 T-cells in protective immunity following infection or vaccination. *Immunology.*

GU, X., ZHOU, F., WANG, Y., FAN, G. & CAO, B. 2020. Respiratory viral sepsis: epidemiology, pathophysiology, diagnosis and treatment. *Eur Respir Rev,* 29.

GUILLEN-GUIO, B., LORENZO-SALAZAR, J. M., MA, S.-F., HOU, P.-C., HERNANDEZ-BEEFTINK, T., CORRALES, A., GARCÍA-LAORDEN, M. I., JOU, J., ESPINOSA, E., MURIEL, A., DOMÍNGUEZ, D., LORENTE, L., MARTÍN, M. M., RODRÍGUEZ-GALLEGO, C., SOLÉ-VIOLÁN, J., AMBRÓS, A., CARRIEDO, D., BLANCO, J., AÑÓN, J. M., REILLY, J. P., JONES, T. K., ITTNER, C. A. G., FENG, R.,

SCHÖNEWECK, F., KIEHNTOPF, M., NOTH, I., SCHOLZ, M., BRUNKHORST, F. M., SCHERAG, A., MEYER, N. J., VILLAR, J. & FLORES, C. 2020. Sepsis-associated acute respiratory distress syndrome in individuals of European ancestry: a genome-wide association study. *The Lancet Respiratory Medicine,* 8**,** 258-266.

GUPTA, N. T., VANDER HEIDEN, J. A., UDUMAN, M., GADALA-MARIA, D., YAARI, G. & KLEINSTEIN, S. H. 2015. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics,* 31**,** 3356-8.

GUTZEIT, C., CHEN, K. & CERUTTI, A. 2018. The enigmatic function of IgD: some answers at last. *Eur J Immunol,* 48**,** 1101-1113.

GYAWALI, B., RAMAKRISHNA, K. & DHAMOON, A. S. 2019. Sepsis: The evolution in definition, pathophysiology, and management. *SAGE Open Med,* 7**,** 2050312119835043.

H. MOSS, P. A. & BELL, J. I. 1996. Comparative sequence analysis of the human T cell receptor TCRA and TCRB CDR3 regions. *Human Immunology,* 48**,** 32-38.

HANSEN, I. S., BAETEN, D. L. P. & DEN DUNNEN, J. 2019. The inflammatory function of human IgA. *Cell Mol Life Sci,* 76**,** 1041-1055.

HAQUE, K. N., ZAIDI, M. H. & BAHAKIM, H. 1988. IgM-Enriched Intravenous Immunoglobulin Therapy in Neonatal Sepsis. *American Journal of Diseases of Children,* 142**,** 1293-1296.

HASE, H., TANI, K., NAGAYAMA, H., WATARI, K., TAKAHASHI, S., OOI, J., SHIRAFUJI, N., ISEKI, T., NAKAZAKI, Y., YAMASHITA, T., NAKAMURA, T., MASUNAGA, A., MAEKAWA, T., TOJO, A. & ASANO, S. 2000. TCR-Vβ repertoire analysis with RT-PCR was useful for the early detection of pulmonary relapsed T-cell lymphoma after autologous peripheral blood stem cell transplantation. *American Journal of Hematology,* 64**,** 124-127.

HE, B., LIU, S., XU, M., HU, Y., LV, K., WANG, Y., MA, Y., ZHAI, Y., YUE, X., LIU, L., LU, H., ZHOU, S., LI, P., MAI, G., HUANG, X., LI, C., CHEN, S., YE, S., ZHAO, P., YANG, Y., LI, X., JIE, Y., SHI, M., YANG, J., SHU, Y. & CHEN, Y. Q. 2022. Comparative global B cell receptor repertoire difference induced by SARS-CoV-2 infection or vaccination via single-cell V(D)J sequencing. *Emerg Microbes Infect,* 11**,** 2007-2020.

HE, W., XIAO, K., FANG, M. & XIE, L. 2021. Immune Cell Number, Phenotype, and Function in the Elderly with Sepsis. *Aging Dis,* 12**,** 277-296.

HERSHBERG, U., MENG, W., ZHANG, B., HAFF, N., ST CLAIR, E. W., COHEN, P. L., MCNAIR, P. D., LI, L., LEVESQUE, M. C. & LUNING PRAK, E. T. 2014. Persistence and selection of an expanded B-cell clone in the setting of rituximab therapy for Sjögren's syndrome. *Arthritis Research & Therapy,* 16**,** R51.

HESSON, L. B., DUNWELL, T. L., COOPER, W. N., CATCHPOOLE, D., BRINI, A. T., CHIARAMONTE, R., GRIFFITHS, M., CHALMERS, A. D., MAHER, E. R. & LATIF, F. 2009. The novel RASSF6 and RASSF10 candidate tumour suppressor genes are frequently epigenetically inactivated in childhood leukaemias. *Mol Cancer,* 8**,** 42.

HINZ, J., BUTTNER, B., KRIESEL, F., STEINAU, M., FREDERIK POPOV, A., GHADIMI, M., BEISSBARTH, T., TZVETKOV, M., BERGMANN, I. & MANSUR, A. 2017. The FER rs4957796 TT genotype is associated with unfavorable 90-day survival in Caucasian patients with severe ARDS due to pneumonia. *Sci Rep,* 7**,** 9887.

HIRSH, M., DYUGOVSKAYA, L., KAPLAN, V. & KRAUSZ, M. M. 2004. Response of lung gammadelta T cells to experimental sepsis in mice. *Immunology,* 112**,** 153-60.

HJELHOLT, A., CHRISTIANSEN, G., SORENSEN, U. S. & BIRKELUND, S. 2013. IgG subclass profiles in normal human sera of antibodies specific to five kinds of microbial antigens. *Pathog Dis,* 67**,** 206-13.

HOCKETT, R. D., DE VILLARTAY, J. P., POLLOCK, K., POPLACK, D. G., COHEN, D. I. & KORSMEYER, S. J. 1988. Human T-cell antigen receptor (TCR) delta-chain locus and elements responsible for its deletion are within the TCR alpha-chain locus. *Proceedings of the National Academy of Sciences,* 85**,** 9694-9698.

HOFFMAN, W., LAKKIS, F. G. & CHALASANI, G. 2016. B Cells, Antibodies, and More. *Clin J Am Soc Nephrol,* 11**,** 137-54.

HORE, V., VINUELA, A., BUIL, A., KNIGHT, J., MCCARTHY, M. I., SMALL, K. & MARCHINI, J. 2016. Tensor decomposition for multiple-tissue gene expression experiments. *Nat Genet,* 48**,** 1094-100.

HOSSAIN, S., IWASA, H., SARKAR, A., MARUYAMA, J., ARIMOTO-MATSUZAKI, K. & HATA, Y. 2018. The RASSF6 Tumor Suppressor Protein Regulates Apoptosis and Cell Cycle Progression via Retinoblastoma Protein. *Mol Cell Biol,* 38.

HOSTE, L., ROELS, L., NAESENS, L., BOSTEELS, V., VANHEE, S., DUPONT, S., BOSTEELS, C., BROWAEYS, R., VANDAMME, N., VERSTAEN, K., ROELS, J., VAN DAMME, K. F. A., MAES, B., DE LEEUW, E., DECLERCQ, J., AEGERTER, H., SEYS, L., SMOLE, U., DE PRIJCK, S., VANHEERSWYNGHELS, M., CLAES, K., DEBACKER, V., VAN ISTERDAEL, G., BACKERS, L., CLAES, K. B. M., BASTARD, P., JOUANGUY, E., ZHANG, S.-Y., METS, G., DEHOORNE, J., VANDEKERCKHOVE, K., SCHELSTRAETE, P., WILLEMS, J., CLINICIANS, M.-C., STORDEUR, P., JANSSENS, S., BEYAERT, R., SAEYS, Y., CASANOVA, J.-L., LAMBRECHT, B. N., HAERYNCK, F. & TAVERNIER, S. J. 2021. TIM3+TRBV11-2 T cells and IFNγ signature in patrolling monocytes and CD16+ NK cells delineate MIS-C. *Journal of Experimental Medicine,* 219.

HOTCHKISS, R. S., COLSTON, E., YENDE, S., ANGUS, D. C., MOLDAWER, L. L., CROUSER, E. D., MARTIN, G. S., COOPERSMITH, C. M., BRAKENRIDGE, S., MAYR, F. B., PARK, P. K., YE, J., CATLETT, I. M., GIRGIS, I. G. & GRASELA, D. M. 2019. Immune Checkpoint Inhibition in Sepsis: A Phase 1b Randomized, Placebo-Controlled, Single Ascending Dose Study of Antiprogrammed Cell Death-Ligand 1 Antibody (BMS-936559). *Crit Care Med,* 47**,** 632-642.

HOTCHKISS, R. S. & KARL, I. E. 2003. The Pathophysiology and Treatment of Sepsis. *New England Journal of Medicine,* 348**,** 138-150.

HOTCHKISS, R. S., MONNERET, G. & PAYEN, D. 2013. Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy. *Nat Rev Immunol,* 13**,** 862-74.

HOU, D., CHEN, C., SEELY, E. J., CHEN, S. & SONG, Y. 2016. High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease. *Frontiers in Immunology,* 7.

HOU, X., ZENG, P., ZHANG, X., CHEN, J., LIANG, Y., YANG, J., YANG, Y., LIU, X. & DIAO, H. 2019a. Shorter TCR β-Chains Are Highly Enriched During Thymic Selection and Antigen-Driven Selection. *Frontiers in Immunology,* 10.

HOU, X. M., BALOCH, Z., ZHENG, Z. H., ZHANG, W. H., FENG, Y., LI, D. D., WU, X. A. & YANG, S. H. 2019b. Knockdown of CHPF suppresses cell progression of non-small-cell lung cancer. *Cancer Manag Res,* 11**,** 3275-3283.

HOVAV, A.-H. 2017. Human γδ T cells: rapid, stable and clonally reactive. *Cellular & Molecular Immunology,* 14**,** 646-648.

HUANG, D., LIU, A. Y. N., LEUNG, K. S. & TANG, N. L. S. 2021. Direct Measurement of B Lymphocyte Gene Expression Biomarkers in Peripheral Blood Transcriptomics Enables Early Prediction of Vaccine Seroconversion. *Genes (Basel),* 12.

HUANG, H., SIKORA, M. J., ISLAM, S., CHOWDHURY, R. R., CHIEN, Y.-H., SCRIBA, T. J., DAVIS, M. M. & STEINMETZ, L. M. 2019. Select sequencing of clonally expanded CD8+ T cells reveals limits to clonal expansion. *Proceedings of the National Academy of Sciences,* 116**,** 8995-9001.

HUNG, C. T., TSAI, Y. W., WU, Y. S., YEH, C. F. & YANG, K. C. 2022. The novel role of ER protein TXNDC5 in the pathogenesis of organ fibrosis: mechanistic insights and therapeutic implications. *J Biomed Sci,* 29**,** 63.

HWANG, J. R., BYEON, Y., KIM, D. & PARK, S. G. 2020. Recent insights of T cell receptor-mediated signaling pathways for T cell activation and development. *Exp Mol Med,* 52**,** 750-761.

IGLESIAS, J., MATAMOROS, N., RAGA, S., FERRER, J. M. & MILA, J. 1999. CD95 expression and function on lymphocyte subpopulations in common variable immunodeficiency (CVID); related to increased apoptosis. *Clin Exp Immunol,* 117**,** 138-46.

ILLUMINA. 2023. *Minimize index hopping in multiplexed runs* [Online]. Available: https://www.illumina.com/techniques/sequencing/ngs-library-prep/multiplexing/index-hopping.html [Accessed April 5, 2023].

IM, S. J., HASHIMOTO, M., GERNER, M. Y., LEE, J., KISSICK, H. T., BURGER, M. C., SHAN, Q., HALE, J. S., LEE, J., NASTI, T. H., SHARPE, A. H., FREEMAN, G. J., GERMAIN, R. N., NAKAYA, H. I., XUE, H. H. & AHMED, R. 2016. Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature,* 537**,** 417-421.

IVANOV, S., PAGET, C. & TROTTEIN, F. 2014. Role of Non-conventional T Lymphocytes in Respiratory Infections: The Case of the Pneumococcus. *PLOS Pathogens,* 10**,** e1004300.

IVERSEN, R., ROY, B., STAMNAES, J., HØYDAHL, L. S., HNIDA, K., NEUMANN, R. S., KORPONAY-SZABÓ, I. R., LUNDIN, K. E. A. & SOLLID, L. M. 2019. Efficient T cell-B cell collaboration guides autoantibody epitope bias and onset of celiac disease. *Proc Natl Acad Sci U S A,* 116**,** 15134-15139.

JACKSON, K. J., GAETA, B., SEWELL, W. & COLLINS, A. M. 2004. Exonuclease activity and P nucleotide addition in the generation of the expressed immunoglobulin repertoire. *BMC Immunol,* 5**,** 19.

JACKSON, K. J., KIDD, M. J., WANG, Y. & COLLINS, A. M. 2013. The shape of the lymphocyte receptor repertoire: lessons from the B cell receptor. *Front Immunol,* 4**,** 263.

JARCZAK, D., KLUGE, S. & NIERHAUS, A. 2021. Sepsis-Pathophysiology and Therapeutic Concepts. *Front Med (Lausanne),* 8**,** 628302.

JENSEN, I. J., SJAASTAD, F. V., GRIFFITH, T. S. & BADOVINAC, V. P. 2018. Sepsis-Induced T Cell Immunoparalysis: The Ins and Outs of Impaired T Cell Immunity. *J Immunol,* 200**,** 1543-1553.

JENSEN, K. D., SU, X., SHIN, S., LI, L., YOUSSEF, S., YAMASAKI, S., STEINMAN, L., SAITO, T., LOCKSLEY, R. M., DAVIS, M. M., BAUMGARTH, N. & CHIEN, Y. H. 2008. Thymic selection determines gammadelta T cell effector fate: antigen-naive cells make interleukin-17 and antigen-experienced cells make interferon gamma. *Immunity,* 29**,** 90-100.

JIANG, A. S., WU, Z., WEI, E. X., NI, H., YOU, B., YANG, T. & JIANG, J. G. 2018. Plasma cell myeloma with dual expression of kappa and lambda light chains. *Int J Clin Exp Pathol,* 11**,** 4718-4723.

JIANG, R., HOEHN, K. B., LEE, C. S., PHAM, M. C., HOMER, R. J., DETTERBECK, F. C., ABAN, I., JACOBSON, L., VINCENT, A., NOWAK, R. J., KAMINSKI, H. J., KLEINSTEIN, S. H. & O'CONNOR, K. C. 2020. Thymus-derived B cell clones persist in the circulation after thymectomy in myasthenia gravis. *Proceedings of the National Academy of Sciences,* 117**,** 30649-30660.

JIN, X., JIA, X. & WANG, Y. 2017. Clinical significance of Vδ1 T cell detection in sepsis patients. *Int J Clin Exp Pathol,* 10**,** 7434-7442.

JING, R., SCARFO, I., NAJIA, M. A., LUMMERTZ DA ROCHA, E., HAN, A., SANBORN, M., BINGHAM, T., KUBACZKA, C., JHA, D. K., FALCHETTI, M., SCHLAEGER, T. M., NORTH, T. E., MAUS, M. V. & DALEY, G. Q. 2022. EZH1 repression generates mature iPSC-derived CAR T cells with enhanced antitumor activity. *Cell Stem Cell,* 29**,** 1181-1196.e6.

JOHNSON, S. A., SEALE, S. L., GITTELMAN, R. M., RYTLEWSKI, J. A., ROBINS, H. S. & FIELDS, P. A. 2021. Impact of HLA type, age and chronic viral infection on peripheral T-cell receptor sharing between unrelated individuals. *PLoS One,* 16**,** e0249484.

JONES, B. G., SEALY, R. E., PENKERT, R. R., SURMAN, S. L., MAUL, R. W., NEALE, G., XU, B., GEARHART, P. J. & HURWITZ, J. L. 2019. Complex sex-biased antibody responses: estrogen receptors bind estrogen response elements centered within immunoglobulin heavy chain gene enhancers. *Int Immunol,* 31**,** 141-156.

JOSSE, J. & HUSSON, F. 2016. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software,* 70.

KANG, H. 2013. The prevention and handling of the missing data. *Korean J Anesthesiol,* 64**,** 402-6.

KAPLINSKY, J., LI, A., SUN, A., COFFRE, M., KORALOV, S. B. & ARNAOUT, R. 2014. Antibody repertoire deep sequencing reveals antigen-independent selection in maturing B cells. *Proceedings of the National Academy of Sciences,* 111**,** E2622-E2629.

KARAS, M. 2022. *Repeated Measures Correlation (rmcorr): when may be not an ideal choice to quantify the association* [Online]. Available: https://martakarass.github.io/post/2022-04-27-rmcorr_vs_lmm/ [Accessed 02/05/2023 2023].

KARLSSON, S., VARPULA, M., RUOKONEN, E., PETTILA, V., PARVIAINEN, I., ALA-KOKKO, T. I., KOLHO, E. & RINTALA, E. M. 2007. Incidence, treatment, and outcome of severe sepsis in ICU-treated adults in Finland: the Finnsepsis study. *Intensive Care Med,* 33**,** 435-43.

KEYT, B. A., BALIGA, R., SINCLAIR, A. M., CARROLL, S. F. & PETERSON, M. S. 2020. Structure, Function, and Therapeutic Use of IgM Antibodies. *Antibodies (Basel),* 9.

KHATRI, I., DIKS, A. M., VAN DEN AKKER, E. B., OOSTEN, L. E. M., ZWAGINGA, J. J., REINDERS, M. J. T., VAN DONGEN, J. J. M. & BERKOWSKA, M. A. 2021. Longitudinal Dynamics of Human B-Cell Response at the Single-Cell Level in Response to Tdap Vaccination. *Vaccines,* 9**,** 1352.

KHODADADI, L., CHENG, Q., RADBRUCH, A. & HIEPE, F. 2019. The Maintenance of Memory Plasma Cells. *Front Immunol,* 10**,** 721.

KIM, H. I. & PARK, S. 2019. Sepsis: Early Recognition and Optimized Treatment. *Tuberc Respir Dis (Seoul),* 82**,** 6-14.

KIM, I., BYUN, S. Y., KIM, S., CHOI, S., NOH, J., CHUNG, J. & KIM, B. G. 2021. Computational analysis of B cell receptor repertoires in COVID-19 patients using deep embedded representations of protein sequences. *bioRxiv***,** 2021.08.02.454701.

KIRBY, A. C., NEWTON, D. J., CARDING, S. R. & KAYE, P. M. 2007. Evidence for the involvement of lung-specific gammadelta T cell subsets in local responses to Streptococcus pneumoniae infection. *Eur J Immunol,* 37**,** 3404-13.

KITAURA, K., SHINI, T., MATSUTANI, T. & SUZUKI, R. 2016. A new high-throughput sequencing method for determining diversity and similarity of T cell receptor (TCR) alpha and beta repertoires and identifying potential new invariant TCR alpha chains. *BMC Immunol,* 17**,** 38.

KITAURA, K., YAMASHITA, H., AYABE, H., SHINI, T., MATSUTANI, T. & SUZUKI, R. 2017. Different Somatic Hypermutation Levels among Antibody Subclasses Disclosed by a New Next-Generation Sequencing-Based Antibody Repertoire Analysis. *Front Immunol,* 8**,** 389.

KLEIN, B., TARTE, K., JOURDAN, M., MATHOUK, K., MOREAUX, J., JOURDAN, E., LEGOUFFE, E., DE VOS, J. & ROSSI, J. F. 2003. Survival and proliferation factors of normal and malignant plasma cells. *Int J Hematol,* 78**,** 106-13.

KLEIN, U., GOOSSENS, T., FISCHER, M., KANZLER, H., BRAEUNINGER, A., RAJEWSKY, K. & KUPPERS, R. 1998a. Somatic hypermutation in normal and transformed human B cells. *Immunol Rev,* 162**,** 261-80.

KLEIN, U., RAJEWSKY, K. & KÜPPERS, R. 1998b. Human Immunoglobulin (Ig)M+IgD+ Peripheral Blood B Cells Expressing the CD27 Cell Surface Antigen Carry Somatically Mutated Variable Region Genes: CD27 as a General Marker for Somatically Mutated (Memory) B Cells. *Journal of Experimental Medicine,* 188**,** 1679-1689.

KNAUS, W. A., DRAPER, E. A., WAGNER, D. P. & ZIMMERMAN, J. E. 1985. APACHE II: a severity of disease classification system. *Crit Care Med,* 13**,** 818-29.

KO, C.-L., LIN, W.-Z., LEE, M.-T., CHANG, Y.-T., LIN, H.-C., WU, Y.-S., LIN, J.-F., PAN, K.-T., CHANG, Y.-C., LEE, K.-H., LEE, Y.-L., HSIEH, T.-T., HUANG, J.-H., WANG, C.-H., YANG, S.-S., CHEN, H.-C. & CHU, C.-M. 2022. Genome-wide association study reveals ethnicity-specific SNPs associated with ankylosing spondylitis in the Taiwanese population. *Journal of Translational Medicine,* 20**,** 589.

KOAY, H. F., GHERARDIN, N. A., XU, C., SENEVIRATNA, R., ZHAO, Z., CHEN, Z., FAIRLIE, D. P., MCCLUSKEY, J., PELLICCI, D. G., ULDRICH, A. P. & GODFREY, D. I. 2019. Diverse MR1-restricted T cells in mice and humans. *Nat Commun,* 10**,** 2243.

KOELSCH, K., ZHENG, N. Y., ZHANG, Q., DUTY, A., HELMS, C., MATHIAS, M. D., JARED, M., SMITH, K., CAPRA, J. D. & WILSON, P. C. 2007. Mature B cells class switched to IgD are autoreactive in healthy individuals. *J Clin Invest,* 117**,** 1558-65.

KONDO, Y., MIYAZATO, A., OKAMOTO, K. & TANAKA, H. 2021. Impact of Sex Differences on Mortality in Patients With Sepsis After Trauma: A Nationwide Cohort Study. *Frontiers in Immunology,* 12.

KOU, Z. C., PUHR, J. S., ROJAS, M., MCCORMACK, W. T., GOODENOW, M. M. & SLEASMAN, J. W. 2000. T-Cell receptor Vbeta repertoire CDR3 length diversity differs within CD45RA and CD45RO T-cell subsets in healthy and human immunodeficiency virus-infected children. *Clin Diagn Lab Immunol,* 7**,** 953-9.

KOVALTSUK, A., KRAWCZYK, K., GALSON, J., KELLY, D., DEANE, C. & TRÜCK, J. 2017. How B-Cell Receptor Repertoire Sequencing Can Be Enriched with Structural Antibody Data. *Frontiers in Immunology,* 8**,** 1753.

KOVALTSUK, A., RAYBOULD, M. I. J., WONG, W. K., MARKS, C., KELM, S., SNOWDEN, J., TRUCK, J. & DEANE, C. M. 2020. Structural diversity of B-cell receptor repertoires along the B-cell differentiation axis in humans and mice. *PLoS Comput Biol,* 16**,** e1007636.

KOWARIK, A. & TEMPL, M. 2016. Imputation with the R Package VIM. *Journal of Statistical Software,* 74.

KRAL, J. B., SCHROTTMAIER, W. C., SALZMANN, M. & ASSINGER, A. 2016. Platelet Interaction with Innate Immune Cells. *Transfus Med Hemother,* 43**,** 78-88.

KRISHNA, C., CHOWELL, D., GÖNEN, M., ELHANATI, Y. & CHAN, T. A. 2020. Genetic and environmental determinants of human TCR repertoire diversity. *Immunity & Ageing,* 17**,** 26.

KRONENBERG, M. & ZAJONC, D. M. 2013. A 'GEM' of a cell. *Nat Immunol,* 14**,** 694-5.

KUMAR, A., ROBERTS, D., WOOD, K. E., LIGHT, B., PARRILLO, J. E., SHARMA, S., SUPPES, R., FEINSTEIN, D., ZANOTTI, S., TAIBERG, L., GURKA, D., KUMAR, A. & CHEANG, M. 2006. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med,* 34**,** 1589-96.

KUMAR BHARATHKAR, S., PARKER, B. W., MALYUTIN, A. G., HALOI, N., HUEY-TUBMAN, K. E., TAJKHORSHID, E. & STADTMUELLER, B. M. 2020. The structures of secretory and dimeric immunoglobulin A. *eLife,* 9**,** e56098.

KUMAR, B. V., CONNORS, T. J. & FARBER, D. L. 2018. Human T Cell Development, Localization, and Function throughout Life. *Immunity,* 48**,** 202-213.

KUMAR, D. & SEHRAWAT, S. 2019. Divergent Effects of a Transient Corticosteroid Therapy on Virus-Specific Quiescent and Effector CD8+ T Cells. *Frontiers in Immunology,* 10.

KURLAND, B. F., JOHNSON, L. L., EGLESTON, B. L. & DIEHR, P. H. 2009. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. *Stat Sci,* 24**,** 211.

KUZNETSOVA, A., BROCKHOFF, P. B. & CHRISTENSEN, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software,* 82**,** 1 - 26.

KWOK, A. J., ALLCOCK, A., FERREIRA, R. C., CANO-GAMEZ, E., SMEE, M., BURNHAM, K. L., ZURKE, Y.-X., NOVAK, A., DARWENT, M., BARON, T., BROWN, C., BEER, S., ESPINOSA, A., PANDURO, T., GEORGIOU, D., MARTINEZ, J., THRAVES, H., PEREZ, E., FERNANDEZ, R., SOBRINO, A., SANCHEZ, V., MAGALLANO, R., DINEEN, K., WILSON, J., MCKECHNIE, S., MENTZER, A. J., MONACO, C., UDALOVA, I. A., HINDS, C. J., TODD, J. A., DAVENPORT, E. E., KNIGHT, J. C. & EMERGENCY MEDICINE RESEARCH, O. 2023. Neutrophils and emergency granulopoiesis drive immune suppression and an extreme response endotype during sepsis. *Nature Immunology,* 24**,** 767-779.

LANGE, A., SUNDEN-CULLBERG, J., MAGNUSON, A. & HULTGREN, O. 2017. Soluble B and T Lymphocyte Attenuator Correlates to Disease Severity in Sepsis and High Levels Are Associated with an Increased Risk of Mortality. *PLoS One,* 12**,** e0169176.

LANGFELDER, P. & HORVATH, S. 2007. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol,* 1**,** 54.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LANGFELDER, P., ZHANG, B. & HORVATH, S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics,* 24**,** 719-20.

LASERSON, U., VIGNEAULT, F., GADALA-MARIA, D., YAARI, G., UDUMAN, M., VANDER HEIDEN, J. A., KELTON, W., TAEK JUNG, S., LIU, Y., LASERSON, J., CHARI, R., LEE, J. H., BACHELET, I., HICKEY, B., LIEBERMAN-AIDEN, E., HANCZARUK, B., SIMEN, B. B., EGHOLM, M., KOLLER, D., GEORGIOU, G., KLEINSTEIN, S. H. & CHURCH, G. M. 2014. High-resolution antibody dynamics of vaccine-induced immune responses. *Proc Natl Acad Sci U S A,* 111**,** 4928-33.

LAUC, G., HUFFMAN, J. E., PUČIĆ, M., ZGAGA, L., ADAMCZYK, B., MUŽINIĆ, A., NOVOKMET, M., POLAŠEK, O., GORNIK, O., KRIŠTIĆ, J., KESER, T., VITART, V., SCHEIJEN, B., UH, H. W., MOLOKHIA, M., PATRICK, A. L., MCKEIGUE, P., KOLČIĆ, I., LUKIĆ, I. K., SWANN, O., VAN LEEUWEN, F. N., RUHAAK, L. R., HOUWING-DUISTERMAAT, J. J., SLAGBOOM, P. E., BEEKMAN, M., DE CRAEN, A. J., DEELDER, A. M., ZENG, Q., WANG, W., HASTIE, N. D., GYLLENSTEN, U., WILSON, J. F., WUHRER, M., WRIGHT, A. F., RUDD, P. M., HAYWARD, C., AULCHENKO, Y., CAMPBELL, H. & RUDAN, I. 2013. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet,* 9**,** e1003225.

LAWAND, M., DECHANET-MERVILLE, J. & DIEU-NOSJEAN, M. C. 2017. Key Features of Gamma-Delta T-Cell Subsets in Human Diseases and Their Immunotherapeutic Implications. *Front Immunol,* 8**,** 761.

LE NOURS, J., GHERARDIN, N. A., RAMARATHINAM, S. H., AWAD, W., WIEDE, F., GULLY, B. S., KHANDOKAR, Y., PRAVEENA, T., WUBBEN, J. M., SANDOW, J. J., WEBB, A. I., VON BORSTEL, A., RICE, M. T., REDMOND, S. J., SENEVIRATNA, R., SANDOVAL-ROMERO, M. L., LI, S., SOUTER, M. N. T., ECKLE, S. B. G., CORBETT, A. J., REID, H. H., LIU, L., FAIRLIE, D. P., GILES, E. M., WESTALL, G. P., TOTHILL, R. W., DAVEY, M. S., BERRY, R., TIGANIS, T., MCCLUSKEY, J., PELLICCI, D. G., PURCELL, A. W., ULDRICH, A. P., GODFREY, D. I. & ROSSJOHN, J. 2019. A class of γδ T cell receptors recognize the underside of the antigen-presenting molecule MR1. *Science,* 366**,** 1522-1527.

LEE, M., PARK, C., WOO, J., KIM, J., KHO, I., NAM, D. H., PARK, W. Y., KIM, Y. S., KONG, D. S., LEE, H. W. & KIM, T. J. 2019. Preferential Infiltration of Unique Vγ9Jγ2-Vδ2 T Cells Into Glioblastoma Multiforme. *Front Immunol,* 10**,** 555.

LEE, S. & LEE, D. K. 2018. What is the proper way to apply the multiple comparison test? *Korean J Anesthesiol,* 71**,** 353-360.

LEEM, J., MITCHELL, L. S., FARMERY, J. H. R., BARTON, J. & GALSON, J. D. 2022. Deciphering the language of antibodies using self-supervised learning. *Patterns,* 3**,** 100513.

LENNOX, S. & BENCH, S. 2022. Blood sampling in adult critical care: a mixed methods study. *International Journal of Orthopaedic and Trauma Nursing,* 45**,** 100923.

LEWIS, A. J., SEYMOUR, C. W. & ROSENGART, M. R. 2016. Current Murine Models of Sepsis. *Surg Infect (Larchmt),* 17**,** 385-93.

LI, H. M., HIROI, T., ZHANG, Y., SHI, A., CHEN, G., DE, S., METTER, E. J., WOOD, W. H., 3RD, SHAROV, A., MILNER, J. D., BECKER, K. G., ZHAN, M. & WENG, N. P. 2016. TCRβ repertoire of CD4+ and CD8+ T cells is distinct in richness, distribution, and CDR3 amino acid composition. *J Leukoc Biol,* 99**,** 505-13.

LI, J., ZHOU, D., QIU, W., SHI, Y., YANG, J. J., CHEN, S., WANG, Q. & PAN, H. 2018. Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Sci Rep,* 8**,** 622.

LI, W. & LIU, J. 2022. The Prognostic and Immunotherapeutic Significance of AHSA1 in Pan-Cancer, and Its Relationship With the Proliferation and Metastasis of Hepatocellular Carcinoma. *Front Immunol,* 13**,** 845585.

LIANG, L., MOORE, B. & SONI, A. 2006. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2017: Statistical Brief #261. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs.* Rockville (MD).

LIAO, W., HUA, Z., LIU, C., LIN, L., CHEN, R. & HOU, B. 2017a. Characterization of T-Dependent and T-Independent B Cell Responses to a Virus-like Particle. *J Immunol,* 198**,** 3846-3856.

LIAO, X.-L., FENG, T., ZHANG, J.-Q., CAO, X., WU, Q.-H., XIE, Z.-C., KANG, Y. & LI, H. 2017b. Phenotypic Changes and Impaired Function of Peripheral γδ T Cells in Patients With Sepsis. *Shock,* 48.

LILES, W. C., KIENER, P. A., LEDBETTER, J. A., ARUFFO, A. & KLEBANOFF, S. J. 1996. Differential expression of Fas (CD95) and Fas ligand on normal human phagocytes: implications for the regulation of apoptosis in neutrophils. *Journal of Experimental Medicine,* 184**,** 429-440.

LIN, G. L., MCGINLEY, J. P., DRYSDALE, S. B. & POLLARD, A. J. 2018. Epidemiology and Immune Pathogenesis of Viral Sepsis. *Front Immunol,* 9**,** 2147.

LIN, Y. H., HUNG, S. J., CHEN, Y. L., LIN, C. H., KUNG, T. F., YEH, Y. C., TSENG, J. T. & LIU, T. 2020. Dissecting efficiency of a 5' rapid amplification of cDNA ends (5'-RACE) approach for profiling T-cell receptor beta repertoire. *PLoS One,* 15**,** e0236366.

LINNANDER, E. L., AYEDUN, A., BOATRIGHT, D., ACKERMAN-BARGER, K., MORGENTHALER, T. I., RAY, N., ROY, B., SIMPSON, S. & CURRY, L. A. 2022. Mitigating structural racism to reduce inequities in sepsis outcomes: a mixed methods, longitudinal intervention study. *BMC Health Services Research,* 22**,** 975.

LINNEMANN, C., HEEMSKERK, B., KVISTBORG, P., KLUIN, R. J., BOLOTIN, D. A., CHEN, X., BRESSER, K., NIEUWLAND, M., SCHOTTE, R., MICHELS, S., GOMEZ-EERLAND, R., JAHN, L., HOMBRINK, P., LEGRAND, N., SHU, C. J., MAMEDOV, I. Z., VELDS, A., BLANK, C. U., HAANEN, J. B., TURCHANINOVA, M. A., KERKHOVEN, R. M., SPITS, H., HADRUP, S. R., HEEMSKERK, M. H., BLANKENSTEIN, T., CHUDAKOV, D. M., BENDLE, G. M. & SCHUMACHER, T. N. 2013. High-throughput identification of antigen-specific TCRs by TCR gene capture. *Nat Med,* 19**,** 1534-41.

LIU, V. X., ESCOBAR, G. J., CHAUDHARY, R. & PRESCOTT, H. C. 2018. Healthcare Utilization and Infection in the Week Prior to Sepsis Hospitalization. *Crit Care Med,* 46**,** 513-516.

LIU, W., WANG, H. & XU, C. 2016a. Antigen Receptor Nanoclusters: Small Units with Big Functions. *Trends Immunol,* 37**,** 680-689.

LIU, X. & WU, J. 2018. History, applications, and challenges of immune repertoire research. *Cell Biology and Toxicology,* 34**,** 441-457.

LIU, X., ZHANG, W., ZENG, X., ZHANG, R., DU, Y., HONG, X., CAO, H., SU, Z., WANG, C., WU, J., NIE, C., XU, X. & KRISTIANSEN, K. 2016b. Systematic Comparative Evaluation of Methods for Investigating the TCRbeta Repertoire. *PLoS One,* 11**,** e0152464.

LIU, Y.-J., DE BOUTEILLER, O., ARPIN, C., BRIÈRE, F., GALIBERT, L., HO, S., MARTINEZ-VALDEZ, H., BANCHEREAU, J. & LEBECQUE, S. 1996. Normal Human IgD+IgM− Germinal Center B Cells Can Express Up to 80 Mutations in the Variable Region of Their IgD Transcripts. *Immunity,* 4**,** 603-613.

LO, K. C., SULLIVAN, E., BANNEN, R. M., JIN, H., ROWE, M., LI, H., PINAPATI, R. S., CARTWRIGHT, A. J., TAN, J. C., PATEL, J., KEYSTONE, E. C. & SIMINOVITCH, K. A. 2020. Comprehensive Profiling of the Rheumatoid Arthritis Antibody Repertoire. *Arthritis Rheumatol,* 72**,** 242-250.

LOMAKIN, Y. A., ZAKHAROVA, M. Y., STEPANOV, A. V., DRONINA, M. A., SMIRNOV, I. V., BOBIK, T. V., PYRKOV, A. Y., TIKUNOVA, N. V., SHARANOVA, S. N., BOITSOV, V. M., VYAZMIN, S. Y., KABILOV, M. R., TUPIKIN, A. E., KRASNOV, A. N., BYKOVA, N. A., MEDVEDEVA, Y. A., FRIDMAN, M. V., FAVOROV, A. V., PONOMARENKO, N. A., DUBINA, M. V., BOYKO, A. N., VLASSOV, V. V., BELOGUROV JR, A. A. & GABIBOV, A. G. 2014. Heavy–light chain interrelations of MS-associated immunoglobulins probed by deep sequencing and rational variation. *Molecular Immunology,* 62**,** 305-314.

LONG, Y., ZHANG, Y., GONG, Y., SUN, R., SU, L., LIN, X., SHEN, A., ZHOU, J., CAIJI, Z., WANG, X., LI, D., WU, H. & TAN, H. 2016. Diagnosis of Sepsis with Cell-free DNA by Next-Generation Sequencing Technology in ICU Patients. *Arch Med Res,* 47**,** 365-371.

LONSDALE, D. O., SHAH, R. V. & LIPMAN, J. 2020. Infection, Sepsis and the Inflammatory Response: Mechanisms and Therapy. *Front Med (Lausanne),* 7**,** 588863.

LU, X., YANG, Y. M. & LU, Y. Q. 2022. Immunosenescence: A Critical Factor Associated With Organ Injury After Sepsis. *Front Immunol,* 13**,** 917293.

LUNING PRAK, E. T., MONESTIER, M. & EISENBERG, R. A. 2011. B cell receptor editing in tolerance and autoimmunity. *Ann N Y Acad Sci,* 1217**,** 96-121.

LY, B. & COTTA, C. V. 2013. The utility of the BIOMED-2 primers in the detection of 2 clonal, B-lymphoproliferative disorders simultaneously involving the same site. *Arch Pathol Lab Med,* 137**,** 1654-9.

MA, C., LIU, H., YANG, S., LI, H., LIAO, X. & KANG, Y. 2022. The emerging roles and therapeutic potential of B cells in sepsis. *Front Pharmacol,* 13**,** 1034667.

MA, L., YANG, L., BIN, S., HE, X., PENG, A., LI, Y., ZHANG, T., SUN, S., MA, R. & YAO, X. 2016. Analyzing the CDR3 Repertoire with respect to TCR—Beta Chain V-D-J and V-J Rearrangements in Peripheral T Cells using HTS. *Scientific Reports,* 6**,** 29544.

MAHMOOD, K., ELDEIRAWI, K. & WAHIDI, M. M. 2012. Association of gender with outcomes in critically ill patients. *Crit Care,* 16**,** R92.

MAMANOVA, L., COFFEY, A. J., SCOTT, C. E., KOZAREWA, I., TURNER, E. H., KUMAR, A., HOWARD, E., SHENDURE, J. & TURNER, D. J. 2010. Target-enrichment strategies for next-generation sequencing. *Nat Methods,* 7**,** 111-8.

MANDRIC, I., ROTMAN, J., YANG, H. T., STRAULI, N., MONTOYA, D. J., VAN DER WEY, W., RONAS, J. R., STATZ, B., YAO, D., PETROVA, V., ZELIKOVSKY, A., SPREAFICO, R., SHIFMAN, S., ZAITLEN, N., ROSSETTI, M., ANSEL, K. M., ESKIN, E. & MANGUL, S. 2020. Profiling immunoglobulin repertoires across multiple human tissues using RNA sequencing. *Nat Commun,* 11**,** 3126.

MANSON, J. J., MAURI, C. & EHRENSTEIN, M. R. 2005. Natural serum IgM maintains immunological homeostasis and prevents autoimmunity. *Springer Semin Immunopathol,* 26**,** 425-32.

MARCOU, Q., MORA, T. & WALCZAK, A. M. 2018. High-throughput immune repertoire analysis with IGoR. *Nature Communications,* 9**,** 561.

MARCOULIDES, K. M. & RAYKOV, T. 2019. Evaluation of Variance Inflation Factors in Regression Models Using Latent Variable Modeling Methods. *Educ Psychol Meas,* 79**,** 874-882.

MARIUZZA, R. A., AGNIHOTRI, P. & ORBAN, J. 2020. The structural basis of T-cell receptor (TCR) activation: An enduring enigma. *J Biol Chem,* 295**,** 914-925.

MARTI-CARVAJAL, A. J., SOLA, I., GLUUD, C., LATHYRIS, D. & CARDONA, A. F. 2012. Human recombinant protein C for severe sepsis and septic shock in adult and paediatric patients. *Cochrane Database Syst Rev,* 12**,** CD004388.

MARTIN, M. D. & BADOVINAC, V. P. 2018. Defining Memory CD8 T Cell. *Front Immunol,* 9**,** 2692.

MARTIN, V., BRYAN WU, Y. C., KIPLING, D. & DUNN-WALTERS, D. 2015. Ageing of the B-cell repertoire. *Philos Trans R Soc Lond B Biol Sci,* 370.

MARTINEZ, F., NOVARINO, J., MEJIA, J. E., FAZILLEAU, N. & ALOULOU, M. 2021a. Ageing of T-dependent B cell responses. *Immunol Lett,* 233**,** 97-103.

MARTINEZ, J.-I., SÁNCHEZ, H.-F., VELANDIA, J.-A., URBINA, Z., FLORIÁN, M.-C., MARTÍNEZ, M.-A., GIAMARELLOS-BOURBOULIS, E. J., PINO-PINZÓN, C.-J., ORTIZ, G. & CELIS, E. 2021b. Treatment with IgM-enriched immunoglobulin in sepsis: a matched case-control analysis. *Journal of Critical Care,* 64**,** 120-124.

MATHEW, N. R., JAYANTHAN, J. K., SMIRNOV, I. V., ROBINSON, J. L., AXELSSON, H., NAKKA, S. S., EMMANOUILIDI, A., CZARNEWSKI, P., YEWDELL, W. T., SCHÖN, K., LEBRERO-FERNÁNDEZ, C., BERNASCONI, V., RODIN, W., HARANDI, A. M., LYCKE, N., BORCHERDING, N., YEWDELL, J. W., GREIFF, V., BEMARK, M. & ANGELETTI, D. 2021. Single-cell BCR and transcriptome analysis after influenza infection reveals spatiotemporal dynamics of antigen-specific B cells. *Cell Rep,* 35**,** 109286.

MATHIAS, B., SZPILA, B. E., MOORE, F. A., EFRON, P. A. & MOLDAWER, L. L. 2015. A Review of GM-CSF Therapy in Sepsis. *Medicine (Baltimore),* 94**,** e2044.

MATSUMOTO, M. L. 2022. Molecular Mechanisms of Multimeric Assembly of IgM and IgA. *Annu Rev Immunol,* 40**,** 221-247.

MATSUTANI, T., YOSHIOKA, T., TSURUTA, Y., IWAGAMI, S. & SUZUKI, R. 1997. Analysis of TCRAV and TCRBV Repertoires in Healthy Individuals by Microplate Hybridization Assay. *Human Immunology,* 56**,** 57-69.

MAYR, F. B., YENDE, S. & ANGUS, D. C. 2014. Epidemiology of severe sepsis. *Virulence,* 5**,** 4-11.

MCCARRON, M. J., PARK, P. W. & FOOKSMAN, D. R. 2017. CD138 mediates selection of mature plasma cells by regulating their survival. *Blood,* 129**,** 2749-2759.

MCGOWAN, J. E., JR., BARNES, M. W. & FINLAND, M. 1975. Bacteremia at Boston City Hospital: Occurrence and mortality during 12 selected years (1935-1972), with special reference to hospital-acquired cases. *J Infect Dis,* 132**,** 316-35.

MCKAY, J. T., EGAN, R. P., YAMMANI, R. D., CHEN, L., SHIN, T., YAGITA, H. & HAAS, K. M. 2015. PD-1 Suppresses Protective Immunity to Streptococcus pneumoniae through a B Cell–Intrinsic Mechanism. *The Journal of Immunology,* 194**,** 2289-2299.

MENG, C., ZELEZNIK, O. A., THALLINGER, G. G., KUSTER, B., GHOLAMI, A. M. & CULHANE, A. C. 2016. Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics,* 17**,** 628-641.

MESIN, L., ERSCHING, J. & VICTORA, G. D. 2016. Germinal Center B Cell Dynamics. *Immunity,* 45**,** 471-482.

MI, Y., BURNHAM, K. L., CHARLES, P. D., HEILIG, R., VENDRELL, I., WHALLEY, J., TORRANCE, H. D., ANTCLIFFE, D. B., MAY, S. M., NEVILLE, M. J., BERRIDGE, G., HUTTON, P., GOH, C., RADHAKRISHNAN, J., NESVIZHSKII, A., YU, F., DAVENPORT, E. E., MCKECHNIE, S., DAVIES, R., O'CALLAGHAN, D. J., PATEL, P., KARPE, F., GORDON, A. C., ACKLAND, G. L., HINDS, C. J., FISCHER, R. & KNIGHT, J. C. 2022. High-throughput mass spectrometry maps the sepsis plasma proteome and differences in response. *medRxiv***,** 2022.08.07.22278495.

MICHAUD, C. M. 2009. Global Burden of Infectious Diseases. *Encyclopedia of Microbiology***,** 444-454.

MIHO, E., ROŠKAR, R., GREIFF, V. & REDDY, S. T. 2019. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nature Communications,* 10**,** 1321.

MINA, M. J., KULA, T., LENG, Y., LI, M., DE VRIES, R. D., KNIP, M., SILJANDER, H., REWERS, M., CHOY, D. F., WILSON, M. S., LARMAN, H. B., NELSON, A. N., GRIFFIN, D. E., DE SWART, R. L. & ELLEDGE, S. J. 2019. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science,* 366**,** 599-606.

MITSUNAGA, E. M. & SNYDER, M. P. 2020. Deep Characterization of the Human Antibody Response to Natural Infection Using Longitudinal Immune Repertoire Sequencing. *Mol Cell Proteomics,* 19**,** 278-293.

MOCKRIDGE, C. I., RAHMAN, A., BUCHAN, S., HAMBLIN, T., ISENBERG, D. A., STEVENSON, F. K. & POTTER, K. N. 2004. Common patterns of B cell perturbation and expanded V4-34 immunoglobulin gene usage in autoimmunity and infection. *Autoimmunity,* 37**,** 9-15.

MOLLOY, E. J., WYNN, J. L., BLISS, J., KOENIG, J. M., KEIJ, F. M., MCGOVERN, M., KUESTER, H., TURNER, M. A., GIANNONI, E., MAZELA, J., DEGTYAREVA, M., STRUNK, T., SIMONS, S. H. P., JANOTA, J., PLOTZ, F. B., VAN DEN HOOGEN, A., DE BOODE, W., SCHLAPBACH, L. J., REISS, I. K. M., ON BEHALF OF THE INFECTION, I. I. & IMMUNISATION SECTION OF THE, E. 2020. Neonatal sepsis: need for consensus definition, collaboration and core outcomes. *Pediatric Research,* 88**,** 2-4.

MONAGHAN, S. F., BANERJEE, D., CHUNG, C. S., LOMAS-NEIRA, J., CYGAN, K. J., RHINE, C. L., FAIRBROTHER, W. G., HEFFERNAN, D. S., LEVY, M. M., CIOFFI, W. G. & AYALA, A. 2018. Changes in the process of alternative RNA splicing results in soluble B and T lymphocyte attenuator with biological and clinical implications in critical illness. *Mol Med,* 24**,** 32.

MONROY-IGLESIAS, M. J., CRESCIOLI, S., BECKMANN, K., LE, N., KARAGIANNIS, S. N., VAN HEMELRIJCK, M. & SANTAOLALLA, A. 2022. Antibodies as biomarkers for cancer risk: a systematic review. *Clin Exp Immunol,* 209**,** 46-63.

MONSERRAT, J., DE PABLO, R., DIAZ-MARTIN, D., RODRIGUEZ-ZAPATA, M., DE LA HERA, A., PRIETO, A. & ALVAREZ-MON, M. 2013. Early alterations of B cells in patients with septic shock. *Crit Care,* 17**,** R105.

MORATH, A. & SCHAMEL, W. W. 2020. alphabeta and gammadelta T cell receptors: Similar but different. *J Leukoc Biol,* 107**,** 1045-1055.

MORBACH, H., EICHHORN, E. M., LIESE, J. G. & GIRSCHICK, H. J. 2010. Reference values for B cell subpopulations from infancy to adulthood. *Clin Exp Immunol,* 162**,** 271-9.

MULDER, D. T., MAHE, E. R., DOWAR, M., HANNA, Y., LI, T., NGUYEN, L. T., BUTLER, M. O., HIRANO, N., DELABIE, J., OHASHI, P. S. & PUGH, T. J. 2018. CapTCR-seq: hybrid capture for T-cell receptor repertoire profiling. *Blood Adv,* 2**,** 3506-3514.

MULITA, F., LIOLIS, E., AKINOSOGLOU, K., TCHABASHVILI, L., MAROULIS, I., KAPLANIS, C., VAILAS, M. & PANOS, G. 2022. Postoperative sepsis after colorectal surgery: a prospective single-center observational study and review of the literature. *Prz Gastroenterol,* 17**,** 47-51.

MURO, R., TAKAYANAGI, H. & NITTA, T. 2019. T cell receptor signaling for gammadeltaT cell development. *Inflamm Regen,* 39**,** 6.

NAKANISHI, K., KUKITA, Y., SEGAWA, H., INOUE, N., OHUE, M. & KATO, K. 2016. Characterization of the T-cell receptor beta chain repertoire in tumor-infiltrating lymphocytes. *Cancer Med,* 5**,** 2513-21.

NAKASONE, C., YAMAMOTO, N., NAKAMATSU, M., KINJO, T., MIYAGI, K., UEZU, K., NAKAMURA, K., HIGA, F., ISHIKAWA, H., O'BRIEN R, L., IKUTA, K., KAKU, M., FUJITA, J. & KAWAKAMI, K. 2007. Accumulation of gamma/delta T cells in the lungs and their roles in neutrophil-mediated host defense against pneumococcal infection. *Microbes Infect,* 9**,** 251-8.

NASIR, N., JAMIL, B., SIDDIQUI, S., TALAT, N., KHAN, F. A. & HUSSAIN, R. 2015. Mortality in Sepsis and its relationship with Gender. *Pak J Med Sci,* 31**,** 1201-6.

NEILSON, H. K., FORTIER, J. H., FINESTONE, P. J., OGILBY, C. M., LIU, R., BRIDGES, E. J. & GARBER, G. E. 2023. Diagnostic Delays in Sepsis: Lessons Learned From a Retrospective Study of Canadian Medico-Legal Claims. *Crit Care Explor,* 5**,** e0841.

NEMAZEE, D. 2017. Mechanisms of central tolerance for B cells. *Nat Rev Immunol,* 17**,** 281-294.

NEMAZEE, D. & WEIGERT, M. 2000. Revising B cell receptors. *The Journal of experimental medicine,* 191**,** 1813-1817.

NGUYEN, T. V., PAWLIKOWSKA, P., FIRLEJ, V., ROSSELLI, F. & AOUFOUCHI, S. 2016. V(D)J recombination process and the Pre-B to immature B-cells transition are altered in Fanca(-/-) mice. *Sci Rep,* 6**,** 36906.

NICOLAS, L., COLS, M., CHOI, J. E., CHAUDHURI, J. & VUONG, B. 2018. Generating and repairing genetically programmed DNA breaks during immunoglobulin class switch recombination. *F1000Res,* 7**,** 458.

NIELSEN, S. C. A., YANG, F., HOH, R. A., JACKSON, K. J. L., ROELTGEN, K., LEE, J. Y., RUSTAGI, A., ROGERS, A. J., POWELL, A. E., KIM, P. S., WANG, T. T., PINSKY, B., BLISH, C. A. & BOYD, S. D. 2020a. B cell clonal expansion and convergent antibody responses to SARS-CoV-2. *Res Sq.*

NIELSEN, S. C. A., YANG, F., JACKSON, K. J. L., HOH, R. A., RÖLTGEN, K., STEVENS, B., LEE, J. Y., RUSTAGI, A., ROGERS, A. J., POWELL, A. E., NAJEEB, J., OTRELO-CARDOSO, A. R., YOST, K. E., DANIEL, B., CHANG, H. Y., SATPATHY, A. T., JARDETZKY, T. S., KIM, P. S., WANG, T. T., PINSKY, B. A., BLISH, C. A. & BOYD, S. D. 2020b. Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2. *bioRxiv.*

NIERHAUS, A., BERLOT, G., KINDGEN-MILLES, D., MÜLLER, E. & GIRARDIS, M. 2020. Best-practice IgM- and IgA-enriched immunoglobulin use in patients with sepsis. *Ann Intensive Care,* 10**,** 132.

NIU, X., LI, S., LI, P., PAN, W., WANG, Q., FENG, Y., MO, X., YAN, Q., YE, X., LUO, J., QU, L., WEBER, D., BYRNE-STEELE, M. L., WANG, Z., YU, F., LI, F., MYERS, R. M., LOTZE, M. T., ZHONG, N., HAN, J. & CHEN, L. 2020a. Longitudinal Analysis of T and B Cell Receptor Repertoire Transcripts Reveal Dynamic Immune Response in COVID-19 Patients. *Frontiers in Immunology,* 11.

NIU, X., YAN, Q., YAO, Z., ZHANG, F., QU, L., WANG, C., WANG, C., LEI, H., CHEN, C., LIANG, R., LUO, J., WANG, Q., ZHAO, L., ZHANG, Y., LUO, K., WANG, L., WU, H., LIU, T., LI, P., ZHENG, Z., TAN, Y. J., FENG, L., ZHANG, Z., HAN, J., ZHANG, F. & CHEN, L. 2020b. Longitudinal analysis of the antibody repertoire of a Zika virus-infected patient revealed dynamic changes in antibody response. *Emerg Microbes Infect,* 9**,** 111-123.

NOURI, N. & KLEINSTEIN, S. H. 2020. Somatic hypermutation analysis for improved identification of B cell clonal families from next-generation sequencing data. *PLoS Comput Biol,* 16**,** e1007977.

ODABASHIAN, M., CARLOTTI, E., ARAF, S., OKOSUN, J., SPADA, F., GRIBBEN, J. G., FORCONI, F., STEVENSON, F. K., CALAMINICI, M. & KRYSOV, S. 2020. IGHV sequencing reveals acquired N-glycosylation sites as a clonal and stable event during follicular lymphoma evolution. *Blood,* 135**,** 834-844.

OHM-LAURSEN, L., MENG, H., CHEN, J., ZHOU, J. Q., CORRIGAN, C. J., GOULD, H. J. & KLEINSTEIN, S. H. 2018. Local Clonal Diversification and Dissemination of B Lymphocytes in the Human Bronchial Mucosa. *Front Immunol,* 9**,** 1976.

OKOYE, I., WANG, L., PALLMER, K., RICHTER, K., ICHIMURA, T., HAAS, R., CROUSE, J., CHOI, O., HEATHCOTE, D., LOVO, E., MAURO, C., ABDI, R., OXENIUS, A., RUTSCHMANN, S. & ASHTON-RICKARDT, P. G. 2015. T cell metabolism. The protein LEM promotes $CD8^+$ T cell immunity through effects on mitochondrial respiration. *Science,* 348**,** 995-1001.

ORANGE, J. S. 2005. Congenital immunodeficiencies and sepsis. *Pediatr Crit Care Med,* 6**,** S99-S107.

OSTROVSKY-BERMAN, M., FRANKEL, B., POLAK, P. & YAARI, G. 2021. Immune2vec: Embedding B/T Cell Receptor Sequences in $\mathbb{R}$ (N) Using Natural Language Processing. *Front Immunol,* 12**,** 680687.

OU, M., ZHENG, F., ZHANG, X., LIU, S., TANG, D., ZHU, P., QIU, J. & DAI, Y. 2018. Integrated analysis of B-cell and T-cell receptors by high-throughput sequencing reveals conserved repertoires in IgA nephropathy. *Mol Med Rep,* 17**,** 7027-7036.

PAVLOVIĆ, M., SCHEFFER, L., MOTWANI, K., KANDURI, C., KOMPOVA, R., VAZOV, N., WAAGAN, K., BERNAL, F. L. M., COSTA, A. A., CORRIE, B., AKBAR, R., AL HAJJ, G. S., BALABAN, G., BRUSKO, T. M., CHERNIGOVSKAYA, M., CHRISTLEY, S., COWELL, L. G., FRANK, R., GRYTTEN, I., GUNDERSEN, S., HAFF, I. H., HOVIG, E., HSIEH, P.-H., KLAMBAUER, G., KUIJJER, M. L., LUND-ANDERSEN, C., MARTINI, A., MINOTTO, T., PENSAR, J., RAND, K., RICCARDI, E., ROBERT, P. A., ROCHA, A., SLABODKIN, A., SNAPKOV, I., SOLLID, L. M., TITOV, D., WEBER, C. R., WIDRICH, M., YAARI, G., GREIFF, V. & SANDVE, G. K. 2021. The immuneML ecosystem for machine learning analysis of adaptive immune receptor repertoires. *Nature Machine Intelligence,* 3**,** 936-944.

PELANDA, R. & TORRES, R. M. 2012. Central B-cell tolerance: where selection begins. *Cold Spring Harb Perspect Biol,* 4**,** a007146.

PENG, J., WANG, P., FANG, H., ZHENG, J., ZHONG, C., YANG, Y. & YU, W. 2021. Weighted Gene Co-Expression Analysis Network-Based Analysis on the Candidate Pathways and Hub Genes in Eggplant Bacterial Wilt-Resistance: A Plant Research Study. *Int J Mol Sci,* 22.

PENG, K., NOWICKI, T. S., CAMPBELL, K., PENG, D., NAGAREDDY, A., HUANG, Y.-N., KARLSBERG, A., MILLER, Z., BRITO, J., PAK, V. M., ABEDALTHAGAFI, M. S., BURKHARDT, A. M., ALACHKAR, H., RIBAS, A. & MANGUL, S. 2022. Rigorous benchmarking of T cell receptor repertoire profiling methods for cancer RNA sequencing. *medRxiv*, 2022.03.31.22273249.

PENG, Q., VIJAYA SATYA, R., LEWIS, M., RANDAD, P. & WANG, Y. 2015. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics,* 16**,** 589.

PENGO, N., SCOLARI, M., OLIVA, L., MILAN, E., MAINOLDI, F., RAIMONDI, A., FAGIOLI, C., MERLINI, A., MARIANI, E., PASQUALETTO, E., ORFANELLI, U., PONZONI, M., SITIA, R., CASOLA, S. & CENCI, S. 2013. Plasma cells require autophagy for sustainable immunoglobulin production. *Nat Immunol,* 14**,** 298-305.

PIEPER, K., GRIMBACHER, B. & EIBEL, H. 2013. B-cell biology and development. *J Allergy Clin Immunol,* 131**,** 959-71.

PILCH, H., HOHN, H., FREITAG, K., NEUKIRCH, C., NECKER, A., HADDAD, P., TANNER, B., KNAPSTEIN, P. G. & MAEURER, M. J. 2002. Improved assessment of T-cell receptor (TCR) VB repertoire in clinical specimens: combination of TCR-CDR3 spectratyping with flow cytometry-based TCR VB frequency analysis. *Clin Diagn Lab Immunol,* 9**,** 257-66.

PILZECKER, B. & JACOBS, H. 2019. Mutating for Good: DNA Damage Responses During Somatic Hypermutation. *Front Immunol,* 10**,** 438.

PISTOIA, V., TUMINO, N., VACCA, P., VENEZIANI, I., MORETTA, A., LOCATELLI, F. & MORETTA, L. 2018. Human gammadelta T-Cells: From Surface Receptors to the Therapy of High-Risk Leukemias. *Front Immunol,* 9**,** 984.

PIZZOLATO, G., KAMINSKI, H., TOSOLINI, M., FRANCHINI, D.-M., PONT, F., MARTINS, F., VALLE, C., LABOURDETTE, D., CADOT, S., QUILLET-MARY, A., POUPOT, M., LAURENT, C., YSEBAERT, L., MERAVIGLIA, S., DIELI, F., MERVILLE, P., MILPIED, P., DÉCHANET-MERVILLE, J. & FOURNIÉ, J.-J. 2019. Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRVδ1 and TCRVδ2 γδ T lymphocytes. *Proceedings of the National Academy of Sciences,* 116**,** 11906-11915.

POLILLI, E., ESPOSITO, J. E., FRATTARI, A., TRAVE, F., SOZIO, F., FERRANDU, G., DI IORIO, G. & PARRUTI, G. 2021. Circulating lymphocyte subsets as promising biomarkers to identify septic patients at higher risk of unfavorable outcome. *BMC Infect Dis,* 21**,** 780.

PORRITT, R. A., PASCHOLD, L., RIVAS, M. N., CHENG, M. H., YONKER, L. M., CHANDNANI, H., LOPEZ, M., SIMNICA, D., SCHULTHEIß, C., SANTISKULVONG, C., VAN EYK, J., MCCORMICK, J. K., FASANO, A., BAHAR, I., BINDER, M. & ARDITI, M. 2021. HLA class I–associated expansion of TRBV11-2 T cells in multisystem inflammatory syndrome in children. *The Journal of Clinical Investigation,* 131.

PORTET, S. 2020. A primer on model selection using the Akaike Information Criterion. *Infect Dis Model,* 5**,** 111-128.

POSADA, D. & BUCKLEY, T. R. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol,* 53**,** 793-808.

PRESCOTT, H. C., OSTERHOLZER, J. J., LANGA, K. M., ANGUS, D. C. & IWASHYNA, T. J. 2016. Late mortality after sepsis: propensity matched cohort study. *Bmj,* 353**,** i2375.

PUEL, A., BASTARD, P., BUSTAMANTE, J. & CASANOVA, J.-L. 2022. Human autoantibodies underlying infectious diseases. *Journal of Experimental Medicine,* 219.

QI, C., WANG, Y., LI, P. & ZHAO, J. 2021. Gamma Delta T Cells and Their Pathogenic Role in Psoriasis. *Front Immunol,* 12**,** 627139.

QUAIL, M. A., SWERDLOW, H. & TURNER, D. J. 2009. Improved protocols for the illumina genome analyzer sequencing system. *Curr Protoc Hum Genet,* Chapter 18**,** Unit 18.2.

RAHMEL, T., SCHMITZ, S., NOWAK, H., SCHEPANEK, K., BERGMANN, L., HALBERSTADT, P., HORTER, S., PETERS, J. & ADAMZIK, M. 2020. Long-term mortality and outcome in hospital survivors of septic shock, sepsis, and severe infections: The importance of aftercare. *PLoS One,* 15**,** e0228952.

RAMPOLDI, F., ULLRICH, L. & PRINZ, I. 2020. Revisiting the Interaction of γδ T-Cells and B-Cells. *Cells,* 9.

RAUTANEN, A., MILLS, T. C., GORDON, A. C., HUTTON, P., STEFFENS, M., NUAMAH, R., CHICHE, J.-D., PARKS, T., CHAPMAN, S. J., DAVENPORT, E. E., ELLIOTT, K. S., BION, J., LICHTNER, P., MEITINGER, T., WIENKER, T. F., CAULFIELD, M. J., MEIN, C., BLOOS, F., BOBEK, I., COTOGNI, P., SRAMEK, V., SARAPUU, S., KOBILAY, M., RANIERI, V. M., RELLO, J., SIRGO, G., WEISS, Y. G., RUSSWURM, S., SCHNEIDER, E. M., REINHART, K., HOLLOWAY, P. A. H., KNIGHT, J. C., GARRARD, C. S., RUSSELL, J. A., WALLEY, K. R., STÜBER, F., HILL, A. V. S. & HINDS, C. J. 2015. Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study. *The Lancet Respiratory Medicine,* 3**,** 53-60.

REINHART, K., DANIELS, R., KISSOON, N., MACHADO, F. R., SCHACHTER, R. D. & FINFER, S. 2017. Recognizing Sepsis as a Global Health Priority - A WHO Resolution. *N Engl J Med,* 377**,** 414-417.

REN, D., REN, C., YAO, R., ZHANG, L., LIANG, X., LI, G., WANG, J., MENG, X., LIU, J., YE, Y., LI, H., WEN, S., CHEN, Y., ZHOU, D., HE, X., LI, X., LAI, K., LI, Y. & GUI, S. 2021. The microbiological diagnostic performance of metagenomic next-generation sequencing in patients with sepsis. *BMC Infect Dis,* 21**,** 1257.

RETH, M., NITSCHKE, L., HIKIDA, M. & KUROSAKI, T. 2015. Structure and Signaling Function of the B-Cell Antigen Receptor and Its Coreceptors. *Molecular Biology of B Cells.*

RHEE, C. & KLOMPAS, M. 2020. Sepsis trends: increasing incidence and decreasing mortality, or changing denominator? *J Thorac Dis,* 12**,** S89-S100.

RICKER, E., MANNI, M., FLORES-CASTRO, D., JENKINS, D., GUPTA, S., RIVERA-CORREA, J., MENG, W., ROSENFIELD, A. M., PANNELLINI, T., CHINENOV, Y., SCULCO, P. K., JESSBERGER, R., PRAK, E. T. L. & PERNIS, A. B. 2021. Sex-specific differences in the function and differentiation of ABCs mark TLR7-driven immunopathogenesis. *bioRxiv*, 2021.01.20.427400.

RIDDELL, N. E. Immune Responses: Primary and Secondary. *Encyclopedia of Life Sciences.*

RIESBECK, K. & NORDSTROM, T. 2006. Structure and immunological action of the human pathogen Moraxella catarrhalis IgD-binding protein. *Crit Rev Immunol,* 26**,** 353-76.

ROBERT, I., DANTZER, F. & REINA-SAN-MARTIN, B. 2009. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *Journal of Experimental Medicine,* 206**,** 1047-1056.

ROCHWERG, B., OCZKOWSKI, S. J., SIEMIENIUK, R. A. C., AGORITSAS, T., BELLEY-COTE, E., D'ARAGON, F., DUAN, E., ENGLISH, S., GOSSACK-KEENAN, K., ALGHUROBA, M., SZCZEKLIK, W., MENON, K., ALHAZZANI, W., SEVRANSKY, J., VANDVIK, P. O., ANNANE, D. & GUYATT, G. 2018. Corticosteroids in Sepsis: An Updated Systematic Review and Meta-Analysis. *Crit Care Med,* 46**,** 1411-1420.

ROCK, E. P., SIBBALD, P. R., DAVIS, M. M. & CHIEN, Y. H. 1994. CDR3 length in antigen-specific immune receptors. *J Exp Med,* 179**,** 323-8.

ROCK, K. L., REITS, E. & NEEFJES, J. 2016. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol,* 37**,** 724-737.

RODRÍGUEZ-GIL, A., RITTER, O., SAUL, V. V., WILHELM, J., YANG, C. Y., GROSSCHEDL, R., IMAI, Y., KUBA, K., KRACHT, M. & SCHMITZ, M. L. 2017. The CCR4-NOT complex contributes to repression of Major Histocompatibility Complex class II transcription. *Sci Rep,* 7**,** 3547.

ROSATI, E., DOWDS, C. M., LIASKOU, E., HENRIKSEN, E. K. K., KARLSEN, T. H. & FRANKE, A. 2017. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol,* 17**,** 61.

ROSIER, F., BRISEBARRE, A., DUPUIS, C., BAAKLINI, S., PUTHIER, D., BRUN, C., PRADEL, L. C., RIHET, P. & PAYEN, D. 2021. Genetic Predisposition to the Mortality in Septic Shock Patients: From GWAS to the Identification of a Regulatory Variant Modulating the Activity of a CISH Enhancer. *Int J Mol Sci,* 22.

ROSSO, F., GIORDANO, A., BARBARISI, M. & BARBARISI, A. 2004. From cell-ECM interactions to tissue engineering. *J Cell Physiol,* 199**,** 174-80.

ROTH, D. B. 2014. V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiol Spectr,* 2.

ROUAUD, P., SAINTAMAND, A., SAAD, F., CARRION, C., LECARDEUR, S., COGNE, M. & DENIZOT, Y. 2014. Elucidation of the enigmatic IgD class-switch recombination via germline deletion of the IgH 3' regulatory region. *J Exp Med,* 211**,** 975-85.

RUDD, K. E., JOHNSON, S. C., AGESA, K. M., SHACKELFORD, K. A., TSOI, D., KIEVLAN, D. R., COLOMBARA, D. V., IKUTA, K. S., KISSOON, N., FINFER, S., FLEISCHMANN-STRUZEK, C., MACHADO, F. R., REINHART, K. K., ROWAN, K., SEYMOUR, C. W., WATSON, R. S., WEST, T. E., MARINHO, F., HAY, S. I., LOZANO, R., LOPEZ, A. D., ANGUS, D. C., MURRAY, C. J. L. & NAGHAVI, M. 2020. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet,* 395**,** 200-211.

SACCO, K., CASTAGNOLI, R., VAKKILAINEN, S., LIU, C., DELMONTE, O. M., OGUZ, C., KAPLAN, I. M., ALEHASHEMI, S., BURBELO, P. D., BHUYAN, F., DE JESUS, A. A., DOBBS, K., ROSEN, L. B., CHENG, A., SHAW, E., VAKKILAINEN, M. S., PALA, F., LACK, J., ZHANG, Y., FINK, D. L., OIKONOMOU, V., SNOW, A. L., DALGARD, C. L., CHEN, J., SELLERS, B. A., MONTEALEGRE SANCHEZ, G. A., BARRON, K., REY-JURADO, E., VIAL, C., POLI, M. C., LICARI, A., MONTAGNA, D., MARSEGLIA, G. L., LICCIARDI, F., RAMENGHI, U., DISCEPOLO, V., LO VECCHIO, A., GUARINO, A., EISENSTEIN, E. M., IMBERTI, L., SOTTINI, A., BIONDI, A., MATÓ, S., GERSTBACHER, D., TRUONG, M., STACK, M. A., MAGLIOCCO, M., BOSTICARDO, M., KAWAI, T., DANIELSON, J. J., HULETT, T., ASKENAZI, M., HU, S., BARNETT, J., CHENG, X., KALADI, K., KURAM, V., MACKEY, J., BANSAL, N. M., MARTINS, A. J., PALTERER, B., MATTHEWS, H., MUDUNURI, U., NAMBIAR, M., OLER, A. J., RASTEGAR, A., SAMUEL, S., SHYU, C., WAINGANKAR, V., WEBER, S., XIRASAGAR, S., ESPINOSA, Y., ASTUDILLO, C., PIÑERA, C., GONZÁLEZ, R., DE FILIPPO, M., VOTTO, M., MONTAGNA, L., COHEN, J. I., SU, H. C., KUHNS, D. B., LIONAKIS, M. S., SNYDER, T. M., HOLLAND, S. M., GOLDBACH-MANSKY, R., TSANG, J. S., NOTARANGELO, L. D., GROUP, N. I. R. T. C., CHILE, M. I. S. C. G. & PAVIA PEDIATRIC, C.-G. 2022. Immunopathological signatures in multisystem inflammatory syndrome in children and pediatric COVID-19. *Nature Medicine,* 28**,** 1050-1062.

SAKR, Y., JASCHINSKI, U., WITTEBOLE, X., SZAKMANY, T., LIPMAN, J., NAMENDYS-SILVA, S. A., MARTIN-LOECHES, I., LEONE, M., LUPU, M. N., VINCENT, J. L. & INVESTIGATORS, I. 2018. Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit. *Open Forum Infect Dis,* 5**,** ofy313.

SANTANA, M. A. & ESQUIVEL-GUADARRAMA, F. 2006. Cell Biology of T Cell Activation and Differentiation.

SCHARPING, N., CAFFERATA, A., HEEG, M., NGUYEN, Q. & GOLDRATH, A. 2021. 660 Targeting T cell fates: converting exhaustion to memory to improve immunotherapeutic responses to cancer. *Journal for ImmunoTherapy of Cancer,* 9**,** A688-A688.

SCHATZ, D. G. & JI, Y. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol,* 11**,** 251-63.

SCHICKEL, J. N., GLAUZY, S., NG, Y. S., CHAMBERLAIN, N., MASSAD, C., ISNARDI, I., KATZ, N., UZEL, G., HOLLAND, S. M., PICARD, C., PUEL, A., CASANOVA, J. L. & MEFFRE, E. 2017. Self-reactive VH4-34-expressing IgG B cells recognize commensal bacteria. *J Exp Med,* 214**,** 1991-2003.

SCHMIDT, M. E. & VARGA, S. M. 2018. The CD8 T Cell Response to Respiratory Virus Infections. *Front Immunol,* 9**,** 678.

SCHROEDER, H. W., JR. & CAVACINI, L. 2010. Structure and function of immunoglobulins. *J Allergy Clin Immunol,* 125**,** S41-52.

SCHULTE, W., BERNHAGEN, J. & BUCALA, R. 2013. Cytokines in sepsis: potent immunoregulators and potential therapeutic targets--an updated view. *Mediators Inflamm,* 2013**,** 165974.

SCHUURMAN, J., PERDOK, G. J., GORTER, A. D. & AALBERSE, R. C. 2001. The inter-heavy chain disulfide bonds of IgG4 are in equilibrium with intra-chain disulfide bonds. *Molecular Immunology,* 38**,** 1-8.

SEIFERT, M., PRZEKOPOWITZ, M., TAUDIEN, S., LOLLIES, A., RONGE, V., DREES, B., LINDEMANN, M., HILLEN, U., ENGLER, H., SINGER, B. B. & KUPPERS, R. 2015. Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc Natl Acad Sci U S A,* 112**,** E546-55.

SEIFERT, M., STEIMLE-GRAUER, S. A., GOOSSENS, T., HANSMANN, M. L., BRAUNINGER, A. & KUPPERS, R. 2009. A model for the development of human IgD-only B cells: Genotypic analyses suggest their generation in superantigen driven immune responses. *Mol Immunol,* 46**,** 630-9.

SELITSKY, S. R., MOSE, L. E., SMITH, C. C., CHAI, S., HOADLEY, K. A., DITTMER, D. P., MOSCHOS, S. J., PARKER, J. S. & VINCENT, B. G. 2019. Prognostic value of B cells in cutaneous melanoma. *Genome Medicine,* 11**,** 36.

SENGUPTA, S. 2019. Gamma-delta T cells in glioblastoma immunotherapy. *Glioma,* 2**,** 30-36.

SEO, M., KIM, K., YOON, J., JEONG, J. Y., LEE, H. J., CHO, S. & KIM, H. 2016. RNA-seq analysis for detecting quantitative trait-associated genes. *Sci Rep,* 6**,** 24375.

SEVY, A. M., SOTO, C., BOMBARDI, R. G., MEILER, J. & CROWE, J. E. 2019. Immune repertoire fingerprinting by principal component analysis reveals shared features in subject groups with common exposures. *BMC Bioinformatics,* 20**,** 629.

SEYLER, S. L., KUMAR, A., THORPE, M. F. & BECKSTEIN, O. 2015. Path Similarity Analysis: A Method for Quantifying Macromolecular Pathways. *PLoS Comput Biol,* 11**,** e1004568.

SHANE, A. L., SÁNCHEZ, P. J. & STOLL, B. J. 2017. Neonatal sepsis. *The Lancet,* 390**,** 1770-1780.

SHANKAR-HARI, M. & RUBENFELD, G. D. 2016. Understanding Long-Term Outcomes Following Sepsis: Implications and Challenges. *Curr Infect Dis Rep,* 18**,** 37.

SHANKAR-HARI, M., SINGER, M. & SPENCER, J. 2017. Can Concurrent Abnormalities in Free Light Chains and Immunoglobulin Concentrations Identify a Target Population for Immunoglobulin Trials in Sepsis?*. *Critical Care Medicine,* 45.

SHANKAR-HARI, M., SPENCER, J., SEWELL, W. A., ROWAN, K. M. & SINGER, M. 2012. Bench-to-bedside review: Immunoglobulin therapy for sepsis - biological plausibility from a critical care perspective. *Critical Care,* 16**,** 206.

SHARON, E., SIBENER, L. V., BATTLE, A., FRASER, H. B., GARCIA, K. C. & PRITCHARD, J. K. 2016. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat Genet,* 48**,** 995-1002.

SHEPHERD, F. R. & MCLAREN, J. E. 2020. T Cell Immunity to Bacterial Pathogens: Mechanisms of Immune Control and Bacterial Evasion. *International Journal of Molecular Sciences* [Online], 21.

SHERWOOD, A. M., DESMARAIS, C., LIVINGSTON, R. J., ANDRIESEN, J., HAUSSLER, M., CARLSON, C. S. & ROBINS, H. 2011. Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after αβ and γδ T cell commitment. *Sci Transl Med,* 3**,** 90ra61.

SHEVYREV, D. & TERESHCHENKO, V. 2019. Treg Heterogeneity, Function, and Homeostasis. *Front Immunol,* 10**,** 3100.

SHI, X., SHAO, T., HUO, F., ZHENG, C., LI, W. & JIANG, Z. 2020. An analysis of abnormalities in the B cell receptor repertoire in patients with systemic sclerosis using high-throughput sequencing. *PeerJ,* 8**,** e8370.

SHIFRUT, E., BARUCH, K., GAL, H., NDIFON, W., DECZKOWSKA, A., SCHWARTZ, M. & FRIEDMAN, N. 2013. CD4+ T Cell-Receptor Repertoire Diversity is Compromised in the Spleen but Not in the Bone Marrow of Aged Mice Due to Private and Sporadic Clonal Expansions. *Frontiers in Immunology,* 4.

SHROCK, E. L., SHROCK, C. L. & ELLEDGE, S. J. 2022. VirScan: High-throughput Profiling of Antiviral Antibody Epitopes. *Bio Protoc,* 12.

SHUGAY, M., BAGAEV, D. V., TURCHANINOVA, M. A., BOLOTIN, D. A., BRITANOVA, O. V., PUTINTSEVA, E. V., POGORELYY, M. V., NAZAROV, V. I., ZVYAGIN, I. V., KIRGIZOVA, V. I., KIRGIZOV, K. I., SKOROBOGATOVA, E. V. & CHUDAKOV, D. M. 2015. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput Biol,* 11**,** e1004503.

SHUGAY, M., BRITANOVA, O. V., MERZLYAK, E. M., TURCHANINOVA, M. A., MAMEDOV, I. Z., TUGANBAEV, T. R., BOLOTIN, D. A., STAROVEROV, D. B., PUTINTSEVA, E. V., PLEVOVA, K., LINNEMANN, C., SHAGIN, D., POSPISILOVA, S., LUKYANOV, S., SCHUMACHER, T. N. & CHUDAKOV, D. M. 2014. Towards error-free profiling of immune repertoires. *Nature Methods,* 11**,** 653-655.

SIGAKIS, M. J. G., JEWELL, E., MAILE, M. D., CINTI, S. K., BATEMAN, B. T. & ENGOREN, M. 2019. Culture-Negative and Culture-Positive Sepsis: A Comparison of Characteristics and Outcomes. *Anesth Analg,* 129**,** 1300-1309.

SINGER, M., DEUTSCHMAN, C. S., SEYMOUR, C. W., SHANKAR-HARI, M., ANNANE, D., BAUER, M., BELLOMO, R., BERNARD, G. R., CHICHE, J. D., COOPERSMITH, C. M., HOTCHKISS, R. S., LEVY, M. M., MARSHALL, J. C., MARTIN, G. S., OPAL, S. M., RUBENFELD, G. D., VAN DER POLL, T., VINCENT, J. L. & ANGUS, D. C. 2016. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA,* 315**,** 801-10.

SINGH, M., ALSALEEM, M. & GRAY, C. P. 2023. Neonatal Sepsis. *StatPearls.* Treasure Island (FL).

SIX, A., MARIOTTI-FERRANDIZ, M. E., CHAARA, W., MAGADAN, S., PHAM, H. P., LEFRANC, M. P., MORA, T., THOMAS-VASLIN, V., WALCZAK, A. M. & BOUDINOT, P. 2013. The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Front Immunol,* 4**,** 413.

SJAASTAD, F. V., CONDOTTA, S. A., KOTOV, J. A., PAPE, K. A., DAIL, C., DANAHY, D. B., KUCABA, T. A., TYGRETT, L. T., MURPHY, K. A., CABRERA-PEREZ, J., WALDSCHMIDT, T. J., BADOVINAC, V. P. & GRIFFITH, T. S. 2018. Polymicrobial Sepsis Chronic Immunoparalysis Is Defined by Diminished Ag-Specific T Cell-Dependent B Cell Responses. *Front Immunol,* 9**,** 2532.

SMAKAJ, E., BABRAK, L., OHLIN, M., SHUGAY, M., BRINEY, B., TOSONI, D., GALLI, C., GROBELSEK, V., D'ANGELO, I., OLSON, B., REDDY, S., GREIFF, V., TRUCK, J., MARQUEZ, S., LEES, W. & MIHO, E. 2020. Benchmarking immunoinformatic tools for the analysis of antibody repertoire sequences. *Bioinformatics,* 36**,** 1731-1739.

SMITH-GARVIN, J. E., KORETZKY, G. A. & JORDAN, M. S. 2009. T cell activation. *Annu Rev Immunol,* 27**,** 591-619.

SNIR, O., MESIN, L., GIDONI, M., LUNDIN, K. E. A., YAARI, G. & SOLLID, L. M. 2015. Analysis of Celiac Disease Autoreactive Gut Plasma Cells and Their Corresponding Memory Compartment in Peripheral Blood Using High-Throughput Sequencing. *The Journal of Immunology,* 194**,** 5703-5712.

SONG, L., COHEN, D., OUYANG, Z., CAO, Y., HU, X. & LIU, X. S. 2021. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nature Methods,* 18**,** 627-630.

SONG, W., LI, D., TAO, L., LUO, Q. & CHEN, L. 2020. Solute carrier transporters: the metabolic gatekeepers of immune cells. *Acta Pharm Sin B,* 10**,** 61-78.

SPITS, H. 2002. Development of alphabeta T cells in the human thymus. *Nat Rev Immunol,* 2**,** 760-72.

STAVNEZER, J., GUIKEMA, J. E. & SCHRADER, C. E. 2008. Mechanism and regulation of class switch recombination. *Annu Rev Immunol,* 26**,** 261-92.

STEARNS-KUROSAWA, D. J., OSUCHOWSKI, M. F., VALENTINE, C., KUROSAWA, S. & REMICK, D. G. 2011. The pathogenesis of sepsis. *Annu Rev Pathol,* 6**,** 19-48.

STEBEGG, M., KUMAR, S. D., SILVA-CAYETANO, A., FONSECA, V. R., LINTERMAN, M. A. & GRACA, L. 2018. Regulation of the Germinal Center Response. *Front Immunol,* 9**,** 2469.

STEEN, C. B., LIU, C. L., ALIZADEH, A. A. & NEWMAN, A. M. 2020. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol,* 2117**,** 135-157.

STEKHOVEN, D. J. & BUHLMANN, P. 2012. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics,* 28**,** 112-8.

STEPHENSON, E., REYNOLDS, G., BOTTING, R. A., CALERO-NIETO, F. J., MORGAN, M. D., TUONG, Z. K., BACH, K., SUNGNAK, W., WORLOCK, K. B., YOSHIDA, M., KUMASAKA, N., KANIA, K., ENGELBERT, J., OLABI, B., SPEGAROVA, J. S., WILSON, N. K., MENDE, N., JARDINE, L., GARDNER, L. C. S., GOH, I., HORSFALL, D., MCGRATH, J., WEBB, S., MATHER, M. W., LINDEBOOM, R. G. H., DANN, E., HUANG, N., POLANSKI, K., PRIGMORE, E., GOTHE, F., SCOTT, J., PAYNE, R. P., BAKER, K. F., HANRATH, A. T., SCHIM VAN DER LOEFF, I. C. D., BARR, A. S., SANCHEZ-GONZALEZ, A., BERGAMASCHI, L., MESCIA, F., BARNES, J. L., KILICH, E., DE WILTON, A., SAIGAL, A., SALEH, A., JANES, S. M., SMITH, C. M., GOPEE, N., WILSON, C., COUPLAND, P., COXHEAD, J. M., KISELEV, V. Y., VAN DONGEN, S., BACARDIT, J., KING, H. W., BAKER, S., BRADLEY, J. R., DOUGAN, G., GOODFELLOW, I. G., GUPTA, R. K., HESS, C., KINGSTON, N., LEHNER, P. J., MATHESON, N. J., OWEHAND, W. H., SAUNDERS, C., SMITH, K. G. C., SUMMERS, C., THAVENTHIRAN, J. E. D., TOSHNER, M., WEEKES, M. P., BUCKE, A., CALDER, J., CANNA, L., DOMINGO, J., ELMER, A., FULLER, S., HARRIS, J., HEWITT, S., KENNET, J., JOSE, S., KOURAMPA, J., MEADOWS, A., O'BRIEN, C., PRICE, J., PUBLICO, C., RASTALL, R., RIBEIRO, C., ROWLANDS, J., RUFFOLO, V., TORDESILLAS, H., BULLMAN, B., DUNMORE, B. J., FAWKE, S., GRÄF, S., HODGSON, J., HUANG, C., HUNTER, K., JONES, E., LEGCHENKO, E., MATARA, C., et al. 2021. Single-cell multi-omics analysis of the immune response in COVID-19. *Nature Medicine,* 27**,** 904-916.

STRITTMATTER-KELLER, U., WALTER, C., RAULD, C., EGLI, N., REGAIRAZ, C., RABE, S., ZENKE, G., CARBALLIDO, J. & SCHWEIGHOFFER, T. 2018. Fingerprints of CD8+ T cells on human pre-plasma and memory B cells. *PLoS One,* 13**,** e0208187.

STUART, P. E., TSOI, L. C., NAIR, R. P., GHOSH, M., KABRA, M., SHAIQ, P. A., RAJA, G. K., QAMAR, R., THELMA, B. K., PATRICK, M. T., PARIHAR, A., SINGH, S., KHANDPUR, S., KUMAR, U., WITTIG, M., DEGENHARDT, F., TEJASVI, T., VOORHEES, J. J., WEIDINGER, S., FRANKE, A., ABECASIS, G. R., SHARMA, V. K. & ELDER, J. T. 2022. Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine mapping in the MHC and genome wide. *Human Genetics and Genomics Advances,* 3**,** 100069.

SUN, A., NOVOBRANTSEVA, T. I., COFFRE, M., HEWITT, S. L., JENSEN, K., SKOK, J. A., RAJEWSKY, K. & KORALOV, S. B. 2015. VH replacement in primary immunoglobulin repertoire diversification. *Proc Natl Acad Sci U S A,* 112**,** E458-66.

SUN, J., SUN, B., GAO, Y., HE, F., YANG, L., WANG, M. & ZHOU, W. 2017. Composition and Variation Analysis of the T Cell Receptor beta-Chain Complementarity Determining Region 3 Repertoire in Neonatal Sepsis. *Scand J Immunol,* 86**,** 418-423.

SUN, W., ZHAO, F., XU, Y., HUANG, K., GUO, X., ZHENG, B., LIU, X., LUO, Z., KONG, Y., XU, M., SCHADENDORF, D. & CHEN, Y. 2020. Chondroitin polymerizing factor (CHPF) promotes development of malignant melanoma through regulation of CDK1. *Cell Death & Disease,* 11**,** 496.

SUN, X., NGUYEN, T., ACHOUR, A., KO, A., CIFELLO, J., LING, C., SHARMA, J., HIROI, T., ZHANG, Y., CHIA, C. W., WOOD, W., III, WU, W. W., ZUKLEY, L., PHUE, J.-N., BECKER, K. G., SHEN, R.-F., FERRUCCI, L. & WENG, N.-P. 2022. Longitudinal analysis reveals age-related changes in the T cell receptor repertoire of human T cell subsets. *The Journal of Clinical Investigation,* 132.

SUTTON, B. J., DAVIES, A. M., BAX, H. J. & KARAGIANNIS, S. N. 2019. IgE Antibodies: From Structure to Function and Clinical Translation. *Antibodies (Basel),* 8.

SUTTON, L. A., KOSTARELI, E., HADZIDIMITRIOU, A., DARZENTAS, N., TSAFTARIS, A., ANAGNOSTOPOULOS, A., ROSENQUIST, R. & STAMATOPOULOS, K. 2009. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen. *Blood,* 114**,** 4460-8.

SWEENEY, T. E., WONG, H. R. & KHATRI, P. 2016. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Science Translational Medicine,* 8**,** 346ra91-346ra91.

TABIBIAN-KEISSAR, H., HAZANOV, L., SCHIBY, G., ROSENTHAL, N., RAKOVSKY, A., MICHAELI, M., SHAHAF, G. L., PICKMAN, Y., ROSENBLATT, K., MELAMED, D., DUNN-WALTERS, D., MEHR, R. & BARSHACK, I. 2016. Aging affects B-cell antigen receptor repertoire diversity in primary and secondary lymphoid tissues. *Eur J Immunol,* 46**,** 480-92.

TACCONE, F. S., STORDEUR, P., DE BACKER, D., CRETEUR, J. & VINCENT, J. L. 2009. Gamma-globulin levels in patients with community-acquired septic shock. *Shock,* 32**,** 379-85.

TAEB, A. M., HOOPER, M. H. & MARIK, P. E. 2017. Sepsis: Current Definition, Pathophysiology, Diagnosis, and Management. *Nutr Clin Pract,* 32**,** 296-308.

TANG, F. & ISHWARAN, H. 2017. Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal,* 10**,** 363-377.

TEMBHARE, P., YUAN, C. M., XI, L., MORRIS, J. C., LIEWEHR, D., VENZON, D., JANIK, J. E., RAFFELD, M. & STETLER-STEVENSON, M. 2011. Flow cytometric immunophenotypic assessment of T-cell clonality by Vbeta repertoire analysis: detection of T-cell clonality at diagnosis and monitoring of minimal residual disease following therapy. *Am J Clin Pathol,* 135**,** 890-900.

TOLAR, P., SOHN, H. W. & PIERCE, S. K. 2005. The initiation of antigen-induced B cell antigen receptor signaling viewed in living cells by fluorescence resonance energy transfer. *Nat Immunol,* 6**,** 1168-76.

TOMINO, A., TSUDA, M., AOKI, R., KAJITA, Y., HASHIBA, M., TERAJIMA, T., KANO, H. & TAKEYAMA, N. 2017. Increased PD-1 Expression and Altered T Cell Repertoire Diversity Predict Mortality in Patients with Septic Shock: A Preliminary Study. *PLoS One,* 12**,** e0169653.

TREANOR, B. 2012. B-cell receptor: from resting state to activate. *Immunology,* 136**,** 21-7.

TREANOR, B., DEPOIL, D., GONZALEZ-GRANJA, A., BARRAL, P., WEBER, M., DUSHEK, O., BRUCKBAUER, A. & BATISTA, F. D. 2010. The membrane skeleton controls diffusion dynamics and signaling through the B cell receptor. *Immunity,* 32**,** 187-99.

TRIDENTE, A., CLARKE, G. M., WALDEN, A., MCKECHNIE, S., HUTTON, P., MILLS, G. H., GORDON, A. C., HOLLOWAY, P. A. H., CHICHE, J. D., BION, J., STUBER, F., GARRARD, C., HINDS, C. J. & GEN, O. I. 2014. Patients with faecal peritonitis admitted to European intensive care units: an epidemiological survey of the GenOSept cohort. *Intensive Care Med,* 40**,** 202-210.

TRUNG, N. T., THAU, N. S., BANG, M. H. & SONG, L. H. 2019. PCR-based Sepsis@Quick test is superior in comparison with blood culture for identification of sepsis-causative pathogens. *Sci Rep,* 9**,** 13663.

TSCHÖP, J., MARTIGNONI, A., GOETZMAN, H. S., CHOI, L. G., WANG, Q., NOEL, J. G., OGLE, C. K., PRITTS, T. A., JOHANNIGMAN, J. A., LENTSCH, A. B. & CALDWELL, C. C. 2008. γδ T cells mitigate the organ injury and mortality of sepsis. *Journal of Leukocyte Biology,* 83**,** 581-588.

VAN DER POLL, T., VAN DE VEERDONK, F. L., SCICLUNA, B. P. & NETEA, M. G. 2017. The immunopathology of sepsis and potential therapeutic targets. *Nat Rev Immunol,* 17**,** 407-420.

VAN DER SLIKKE, E. C., AN, A. Y., HANCOCK, R. E. W. & BOUMA, H. R. 2020. Exploring the pathophysiology of post-sepsis syndrome to identify therapeutic opportunities. *eBioMedicine,* 61.

VAN DER VALK, T., VEZZI, F., ORMESTAD, M., DALEN, L. & GUSCHANSKI, K. 2020. Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Mol Ecol Resour,* 20**,** 1171-1181.

VAN DONGEN, J. J., LANGERAK, A. W., BRUGGEMANN, M., EVANS, P. A., HUMMEL, M., LAVENDER, F. L., DELABESSE, E., DAVI, F., SCHUURING, E., GARCIA-SANZ, R., VAN KRIEKEN, J. H., DROESE, J., GONZALEZ, D., BASTARD, C., WHITE, H. E., SPAARGAREN, M., GONZALEZ, M., PARREIRA, A., SMITH, J. L., MORGAN, G. J., KNEBA, M. & MACINTYRE, E. A. 2003. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia,* 17**,** 2257-317.

VANDER HEIDEN, J. A., MARQUEZ, S., MARTHANDAN, N., BUKHARI, S. A. C., BUSSE, C. E., CORRIE, B., HERSHBERG, U., KLEINSTEIN, S. H., MATSEN IV, F. A., RALPH, D. K., ROSENFELD, A. M., SCHRAMM, C. A., CHRISTLEY, S. & LASERSON, U. 2018. AIRR Community Standardized Representations for Annotated Immune Repertoires. *Front Immunol,* 9**,** 2206.

VANDER HEIDEN, J. A., YAARI, G., UDUMAN, M., STERN, J. N., O'CONNOR, K. C., HAFLER, D. A., VIGNEAULT, F. & KLEINSTEIN, S. H. 2014. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics,* 30**,** 1930-2.

VENET, F., BOHÉ, J., DEBARD, A. L., BIENVENU, J., LEPAPE, A. & MONNERET, G. 2005. Both percentage of gammadelta T lymphocytes and CD3 expression are reduced during septic shock. *Crit Care Med,* 33**,** 2836-40.

VENET, F., FILIPE-SANTOS, O., LEPAPE, A., MALCUS, C., POITEVIN-LATER, F., GRIVES, A., PLANTIER, N., PASQUAL, N. & MONNERET, G. 2013. Decreased T-cell repertoire diversity in sepsis: a preliminary study. *Crit Care Med,* 41**,** 111-9.

VENET, F., GEBEILE, R., BANCEL, J., GUIGNANT, C., POITEVIN-LATER, F., MALCUS, C., LEPAPE, A. & MONNERET, G. 2011. Assessment of plasmatic immunoglobulin G, A and M levels in septic shock patients. *Int Immunopharmacol,* 11**,** 2086-90.

VETTERMANN, C. & SCHLISSEL, M. S. 2010. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol Rev,* 237**,** 22-42.

VIDARSSON, G., DEKKERS, G. & RISPENS, T. 2014. IgG subclasses and allotypes: from structure to effector functions. *Front Immunol,* 5**,** 520.

VINCENT, J.-L., RELLO, J., MARSHALL, J., SILVA, E., ANZUETO, A., MARTIN, C. D., MORENO, R., LIPMAN, J., GOMERSALL, C., SAKR, Y., REINHART, K. & EPIC II GROUP OF INVESTIGATORS, F. T. 2009. International Study of the Prevalence and Outcomes of Infection in Intensive Care Units. *JAMA,* 302**,** 2323-2329.

VINCENT, J. L., MORENO, R., TAKALA, J., WILLATTS, S., DE MENDONÇA, A., BRUINING, H., REINHART, C. K., SUTER, P. M. & THIJS, L. G. 1996. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med,* 22**,** 707-10.

VINCENT, J. L., SAKR, Y., SPRUNG, C. L., RANIERI, V. M., REINHART, K., GERLACH, H., MORENO, R., CARLET, J., LE GALL, J. R., PAYEN, D. & SEPSIS OCCURRENCE IN ACUTELY ILL PATIENTS, I. 2006. Sepsis in European intensive care units: results of the SOAP study. *Crit Care Med,* 34**,** 344-53.

VODNALA, S. K., EIL, R., KISHTON, R. J., SUKUMAR, M., YAMAMOTO, T. N., HA, N.-H., LEE, P.-H., SHIN, M., PATEL, S. J., YU, Z., PALMER, D. C., KRUHLAK, M. J., LIU, X., LOCASALE, J. W., HUANG, J., ROYCHOUDHURI, R., FINKEL, T., KLEBANOFF, C. A. & RESTIFO, N. P. 2019. T cell stemness and dysfunction in tumors are triggered by a common mechanism. *Science,* 363**,** eaau0135.

VOS, Q., LEES, A., WU, Z. Q., SNAPPER, C. M. & MOND, J. J. 2000. B-cell activation by T-cell-independent type 2 antigens as an integral part of the humoral immune response to pathogenic microorganisms. *Immunol Rev,* 176**,** 154-70.

WAGENMAKERS, E.-J. & FARRELL, S. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review,* 11**,** 192-196.

WALDEN, A. P., CLARKE, G. M., MCKECHNIE, S., HUTTON, P., GORDON, A. C., RELLO, J., CHICHE, J. D., STUEBER, F., GARRARD, C. S. & HINDS, C. J. 2014. Patients with community acquired pneumonia admitted to European intensive care units: an epidemiological survey of the GenOSept cohort. *Crit Care,* 18**,** R58.

WALSH, E. S., TOLLISON, T. S., BROCHU, H. N., SHAW, B. I., DIVELEY, K. R., CHOU, H., LAW, L., KIRK, A. D., GALE, M. & PENG, X. 2021. Single cell based high-throughput Ig and TCR repertoire sequencing analysis in rhesus macaques. *bioRxiv***,** 2021.08.17.456682.

WALTARI, E., MCGEEVER, A., FRIEDLAND, N., KIM, P. S. & MCCUTCHEON, K. M. 2019. Functional Enrichment and Analysis of Antigen-Specific Memory B Cell Antibody Repertoires in PBMCs. *Front Immunol,* 10**,** 1452.

WANG, C. Y., FANG, Y. X., CHEN, G. H., JIA, H. J., ZENG, S., HE, X. B., FENG, Y., LI, S. J., JIN, Q. W., CHENG, W. Y. & JING, Z. Z. 2017. Analysis of the CDR3 length repertoire and the diversity of T cell receptor α and β chains in swine CD4+ and CD8+ T lymphocytes. *Mol Med Rep,* 16**,** 75-86.

WANG, X., LI, W., ZHU, D., ZHAO, H., CHEN, P. & CHEN, X. 2020a. Characterization of human peripheral blood γδ T cells in patients with sepsis. *Exp Ther Med,* 19**,** 3698-3706.

WANG, Y., LIU, J., BURROWS, P. D. & WANG, J.-Y. 2020b. B Cell Development and Maturation. *In:* WANG, J.-Y. (ed.) *B Cells in Immunity and Tolerance.* Singapore: Springer Singapore.

WEISER, J. N., FERREIRA, D. M. & PATON, J. C. 2018. Streptococcus pneumoniae: transmission, colonization and invasion. *Nat Rev Microbiol,* 16**,** 355-367.

WEKSLER, M. E. & SZABO, P. 2000. The Effect of Age on the B-Cell Repertoire. *Journal of Clinical Immunology,* 20**,** 240-249.

WELEDJI, E. P. & NGOWE, M. N. 2013. The challenge of intra-abdominal sepsis. *International Journal of Surgery,* 11**,** 290-295.

WELLINGHAUSEN, N., WIRTHS, B., ESSIG, A. & WASSILL, L. 2004. Evaluation of the Hyplex BloodScreen Multiplex PCR-Enzyme-linked immunosorbent assay system for direct identification of gram-positive cocci and gram-negative bacilli from positive blood cultures. *J Clin Microbiol,* 42**,** 3147-52.

WEST, M. A. & HEAGY, W. 2002. Endotoxin tolerance: A review. *Critical Care Medicine,* 30.

WESTH, H., LISBY, G., BREYSSE, F., BODDINGHAUS, B., CHOMARAT, M., GANT, V., GOGLIO, A., RAGLIO, A., SCHUSTER, H., STUBER, F., WISSING, H. & HOEFT, A. 2009. Multiplex real-time PCR and blood culture for identification of bloodstream pathogens in patients with suspected sepsis. *Clin Microbiol Infect,* 15**,** 544-51.

WHEELER, D. S., WONG, H. R. & ZINGARELLI, B. 2011. Pediatric Sepsis - Part I: "Children are not small adults!". *Open Inflamm J,* 4**,** 4-15.

WIERSINGA, W. J., LEOPOLD, S. J., CRANENDONK, D. R. & VAN DER POLL, T. 2014. Host innate immune responses to sepsis. *Virulence,* 5**,** 36-44.

WIERSINGA, W. J. & VAN DER POLL, T. 2022. Immunopathophysiology of human sepsis. *EBioMedicine,* 86**,** 104363.

WILLCOX, C. R., VANTOUROUT, P., SALIM, M., ZLATAREVA, I., MELANDRI, D., ZANARDO, L., GEORGE, R., KJAER, S., JEEVES, M., MOHAMMED, F., HAYDAY, A. C. & WILLCOX, B. E. 2019. Butyrophilin-like 3 Directly Binds a Human Vγ4+ T Cell Receptor Using a Modality Distinct from Clonally-Restricted Antigen. *Immunity,* 51**,** 813-825.e4.

WILMORE, J. R., GAUDETTE, B. T., GOMEZ ATRIA, D., HASHEMI, T., JONES, D. D., GARDNER, C. A., COLE, S. D., MISIC, A. M., BEITING, D. P. & ALLMAN, D. 2018. Commensal Microbes Induce Serum IgA Responses that Protect against Polymicrobial Sepsis. *Cell Host Microbe,* 23**,** 302-311 e3.

WOOF, J. M. & KERR, M. A. 2006. The function of immunoglobulin A in immunity. *J Pathol,* 208**,** 270-82.

WU, Y., BISWAS, D., USAITE, I., ANGELOVA, M., BOEING, S., KARASAKI, T., VEERIAH, S., CZYZEWSKA-KHAN, J., MORTON, C., JOSEPH, M., HESSEY, S., READING, J., GEORGIOU, A., AL-BAKIR, M., BIRKBAK, N. J., PRICE, G., KHALIL, M., KERR, K., RICHARDSON, S., CHEYNE, H., CRUICKSHANK, T., WILSON, G. A., ROSENTHAL, R., AERTS, H., HEWISH, M., ANAND, G., KHAN, S., LAU, K., SHEAFF, M., SCHMID, P., LIM, L., CONIBEAR, J., SCHWARZ, R., KAUFMANN, T. L., HUSKA, M., SHAW, J., RILEY, J., PRIMROSE, L., FENNELL, D., HACKSHAW, A., NGAI, Y., SHARP, A., PRESSEY, O., SMITH, S., GOWER, N., DHANDA, H. K., CHAN, K., CHAKRABORTY, S., LITCHFIELD, K., THAKKAR, K., TUGWOOD, J., CLIPSON, A., DIVE, C., ROTHWELL, D., KERR, A., KILGOUR, E., MORGAN, F., KORNASZEWSKA, M., ATTANOOS, R., DAVIES, H., BAKER, K., CARTER, M., LINDSAY, C. R., GOMES, F., BLACKHALL, F., PRIEST, L., KREBS, M. G., CHATURVEDI, A., OLIVEIRA, P., SZALLASI, Z., ROYLE, G., VEIGA, C., SKRZYPSKI, M., SALGADO, R., DIOSSY, M., KIRK, A., ASIF, M., BUTLER, J., BILANCIA, R., KOSTOULAS, N., THOMAS, M., MACKENZIE, M., WILCOX, M., NAKAS, A., RATHINAM, S., BOYLES, R., TUFAIL, M., BAJAJ, A., ANG, K., CHOWDHRY, M. F., SHACKCLOTH, M., ASANTE-SIAW, J., LEEK, A., TOTTEN, N., HODGKINSON, J. D., VAN LOO, P., MONTEIRO, W., MARSHAL, H., BLYTH, K. G., DICK, C., et al. 2022a. A local human Vδ1 T cell population is associated with survival in nonsmall-cell lung cancer. *Nature Cancer,* 3**,** 696-709.

WU, Y., WANG, Y., SHI, H., HU, H., YI, L. & HOU, J. 2022b. Time-course transcriptome and WGCNA analysis revealed the drought response mechanism of two sunflower inbred lines. *PLoS One,* 17**,** e0265447.

XU-MONETTE, Z. Y., LI, J., XIA, Y., CROSSLEY, B., BREMEL, R. D., MIAO, Y., XIAO, M., SNYDER, T., MANYAM, G. C., TAN, X., ZHANG, H., VISCO, C., TZANKOV, A., DYBKAER, K., BHAGAT, G., TAM, W., YOU, H., HSI, E. D., VAN KRIEKEN, J. H., HUH, J., PONZONI, M., FERRERI, A. J. M., MØLLER, M. B., PIRIS, M. A., WINTER, J. N., MEDEIROS, J. T., XU, B., LI, Y., KIRSCH, I. & YOUNG, K. H. 2019. Immunoglobulin somatic hypermutation has clinical impact in DLBCL and potential implications for immune checkpoint blockade and neoantigen-based immunotherapies. *Journal for ImmunoTherapy of Cancer,* 7**,** 272.

XU, D. 2006a. Dual Surface Immunoglobulin Light-Chain Expression in B-Cell Lymphoproliferative Disorders. *Archives of Pathology & Laboratory Medicine,* 130**,** 853-856.

XU, D. 2006b. Dual surface immunoglobulin light-chain expression in B-cell lymphoproliferative disorders. *Arch Pathol Lab Med,* 130**,** 853-6.

XU, W., HE, B., CHIU, A., CHADBURN, A., SHAN, M., BULDYS, M., DING, A., KNOWLES, D. M., SANTINI, P. A. & CERUTTI, A. 2007. Epithelial cells trigger frontline immunoglobulin class switching through a pathway regulated by the inhibitor SLPI. *Nature Immunology,* 8**,** 294-303.

XU, W., LAU, Z. W. X., FULOP, T. & LARBI, A. 2020a. The Aging of gammadelta T Cells. *Cells,* 9.

XU, W., LAU, Z. W. X., FULOP, T. & LARBI, A. 2020b. The Aging of γδ T Cells. *Cells,* 9.

XU, Z., ZAN, H., PONE, E. J., MAI, T. & CASALI, P. 2012. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol,* 12**,** 517-31.

YAARI, G. & KLEINSTEIN, S. H. 2015. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med,* 7**,** 121.

YANG, J. & RETH, M. 2010. Oligomeric organization of the B-cell antigen receptor on resting cells. *Nature,* 467**,** 465-9.

YANG, X. W., LI, H., FENG, T., ZHANG, W., SONG, X. R., MA, C. Y., NIE, M., WANG, L., TAN, X., KANG, Y. & LIAO, X. 2021. Impairment of antigen-presenting function of peripheral γδ T cells in patients with sepsis. *Clin Exp Immunol.*

YE, J., MA, N., MADDEN, T. L. & OSTELL, J. M. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res,* 41**,** W34-40.

YE, Z., ZHANG, Y., HUANG, N., CHEN, S., WU, X. & LI, L. 2022. Immune repertoire and evolutionary trajectory analysis in the development of diabetic nephropathy. *Front Immunol,* 13**,** 1006137.

YOSHIDA, K., COLOGNE, J. B., CORDOVA, K., MISUMI, M., YAMAOKA, M., KYOIZUMI, S., HAYASHI, T., ROBINS, H. & KUSUNOKI, Y. 2017. Aging-related changes in human T-cell repertoire over 20years delineated by deep sequencing of peripheral T-cell receptors. *Exp Gerontol,* 96**,** 29-37.

YOUNESIAN, S., SHAHKARAMI, S., GHAFFARI, P., ALIZADEH, S., MEHRASA, R., GHAVAMZADEH, A. & GHAFFARI, S. H. 2017. DNA hypermethylation of tumor suppressor genes RASSF6 and RASSF10 as independent prognostic factors in adult acute lymphoblastic leukemia. *Leuk Res,* 61**,** 33-38.

YU, K., SHI, J., LU, D. & YANG, Q. 2019. Comparative analysis of CDR3 regions in paired human αβ CD8 T cells. *FEBS Open Bio,* 9**,** 1450-1459.

ZBOROMYRSKA, Y., CILLONIZ, C., COBOS-TRIGUEROS, N., ALMELA, M., HURTADO, J. C., VERGARA, A., MATA, C., SORIANO, A., MENSA, J., MARCO, F. & VILA, J. 2019. Evaluation of the Magicplex Sepsis Real-Time Test for the Rapid Diagnosis of Bloodstream Infections in Adults. *Front Cell Infect Microbiol,* 9**,** 56.

ZELAZOWSKA, M. A., DONG, Q., PLUMMER, J. B., ZHONG, Y., LIU, B., KRUG, L. T. & MCBRIDE, K. M. 2020. Gammaherpesvirus-infected germinal center cells express a distinct immunoglobulin repertoire. *Life Sci Alliance,* 3.

ZELLWEGER, R., WICHMANN, M. W., AYALA, A., STEIN, S., DEMASO, C. M. & CHAUDRY, I. H. 1997. Females in proestrus state maintain splenic immune functions and tolerate sepsis better than males. *Crit Care Med,* 25**,** 106-10.

ZEMLIN, M., SCHELONKA, R. L., BAUER, K. & SCHROEDER, H. W. 2002. Regulation and chance in the ontogeny of B and T cell antigen receptor repertoires. *Immunologic Research,* 26**,** 265-278.

ZENZ, R., EFERL, R., SCHEINECKER, C., REDLICH, K., SMOLEN, J., SCHONTHALER, H. B., KENNER, L., TSCHACHLER, E. & WAGNER, E. F. 2008. Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease. *Arthritis Research & Therapy,* 10**,** 201.

ZHANG, Y., YAN, Q., LUO, K., HE, P., HOU, R., ZHAO, X., WANG, Q., YI, H., LIANG, H., DENG, Y., HU, F., LI, F., LIU, X., FENG, Y., LI, P., QU, L., CHEN, Z., PAN-HAMMARSTRÖM, Q., FENG, L., NIU, X. & CHEN, L. 2022. Analysis of B Cell Receptor Repertoires Reveals Key Signatures of the Systemic B Cell Response after SARS-CoV-2 Infection. *J Virol,* 96**,** e0160021.

ZHAO, Y., NIU, C. & CUI, J. 2018. Gamma-delta (gammadelta) T cells: friend or foe in cancer development? *J Transl Med,* 16**,** 3.

ZHENG, B., YANG, Y., CHEN, L., WU, M. & ZHOU, S. 2022. B-cell receptor repertoire sequencing: Deeper digging into the mechanisms and clinical aspects of immune-mediated diseases. *iScience,* 25**,** 105002.

ZHI, H., DAI, J., LIU, J., ZHU, J., NEWMAN, D. K., GAO, C. & NEWMAN, P. J. 2015. Platelet Activation and Thrombus Formation over IgG Immune Complexes Requires Integrin alphaIIbbeta3 and Lyn Kinase. *PLoS One,* 10**,** e0135738.

ZHOU, J., LOTTENBACH, K. R., BARENKAMP, S. J. & REASON, D. C. 2004. Somatic hypermutation and diverse immunoglobulin gene usage in the human antibody response to the capsular polysaccharide of Streptococcus pneumoniae Type 6B. *Infect Immun,* 72**,** 3505-14.

ZHU, D., LOSSOS, C., CHAPMAN-FREDRICKS, J. R., MATTHEWS, J. M., IKPATT, O. F., RUIZ, P. & LOSSOS, I. S. 2011. Biased use of the IGHV4 family and evidence for antigen selection in Chlamydophila psittaci-negative ocular adnexal extranodal marginal zone lymphomas. *PLoS One,* 6**,** e29114.

ZIEGLER, E. J., FISHER, C. J., JR., SPRUNG, C. L., STRAUBE, R. C., SADOFF, J. C., FOULKE, G. E., WORTEL, C. H., FINK, M. P., DELLINGER, R. P., TENG, N. N. & ET AL. 1991. Treatment of gram-negative bacteremia and septic shock with HA-1A human monoclonal antibody against endotoxin. A randomized, double-blind, placebo-controlled trial. The HA-1A Sepsis Study Group. *N Engl J Med,* 324**,** 429-36.

ZIEGLER, H., WELKER, C., STERK, M., HAARER, J., RAMMENSEE, H.-G., HANDGRETINGER, R. & SCHILBACH, K. 2014. Human peripheral CD4+ Vdelta1+ gamma delta T cells can develop into alpha beta T cells. *Frontiers in Immunology,* 5.